




INCASI Working Paper Series

2019, No. 7

 **INCASI** *International Network for
Comparative Analysis of Social Inequalities*



Vecino más cercano en el espacio social: un método de imputación factorial

Sandra Fachelli
Pedro López-Roldán



European
Commission

Horizon 2020
European Union funding
for Research & Innovation

Marie Skłodowska-Curie Actions (MSCA)
Research and Innovation Staff Exchange (RISE)

H2020-MSCA-RISE-2015
GA-691004



Vecino más cercano en el espacio social: un método de imputación factorial

Sandra Fachelli¹
Pedro López-Roldán²

¹ Universidad Pablo de Olavide
Grup de Recerca en Educació i Treball
Universitat Autònoma de Barcelona, Spain
Sandra.Fachelli@uab.es

² Centre d'Estudis Sociològics sobre la Vida Quotidiana i el Treball
Institut d'Estudis del Treball
Universitat Autònoma de Barcelona, Spain
Pedro.Lopez.Roldan@uab.es

INCASI Working Paper Series is an online publication under *Creative Commons* license. Any person is free to copy, distribute or publicly communicate the work, according to the following conditions:



Attribution. All CC licenses require that others who use your work in any way must give you credit the way you request, but not in a way that suggests you endorse them or their use. If they want to use your work without giving you credit or for endorsement purposes, they must get your permission first.



NonCommercial. You let others copy, distribute, display, perform, and (unless you have chosen NoDerivatives) modify and use your work for any purpose other than commercially unless they get your permission first.



NoDerivatives. You let others copy, distribute, display and perform only original copies of your work. If they want to modify your work, they must get your permission first.

There are no additional restrictions. You cannot apply legal terms or technological measures that legally restrict doing what the license allows.

This working paper was elaborated in the context of INCASI Network, a European project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie GA, No. 691004, and coordinated by Dr. Pedro López-Roldán. This article reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

Digital edition: <https://ddd.uab.cat/record/204787>

Dipòsit Digital de Documents
Bellaterra, Cerdantola del Vallès (Barcelona)
Universitat Autònoma de Barcelona



Vecino más cercano en el espacio social: un método de imputación factorial¹

Sandra Fachelli
Pedro López-Roldán

Resumen

En este artículo nos proponemos establecer un proceso metodológico general, operativo y técnico para realizar imputación de valores faltantes, en este caso sobre la variable de ingresos. El modelo de análisis que se utiliza de referencia para la imputación está conformado por aspectos vinculados al ámbito laboral como al sociodemográfico. La metodología utilizada para la imputación es el Análisis Factorial de Correspondencias Múltiples, por tanto, es un análisis multivariable de interdependencia que permite conformar un espacio social donde se inscribe la población trabajadora. A partir de calcular las puntuaciones factoriales en ese espacio, los vecinos más cercanos, determinados en términos de distancia euclidiana, esto es, de distancia social, se convertirán en donantes de sus ingresos a aquellos que son próximos según un mismo perfil social, pero de los que no se dispone de la información de la variable.

Palabras clave

Imputación, ingresos, vecino más cercano, espacio social, distancia social, análisis factorial de correspondencia múltiples

Índice

1. Presentación. 2. Antecedentes. 3. Procedimientos de imputación. 4. Modelo de análisis y Metodología. 5. Datos. 6. Resultados. 6.1. Especificación del modelo teórico-empírico. 6.2. Procedimiento técnico de imputación. 6.2.1. Análisis de Correspondencias múltiples. Determinación del número de factores e interpretación. Cálculo de las puntuaciones factoriales. 6.2.2. Cálculo de la matriz de distancias. 6.2.3. Determinación del vecino más cercano. 6.2.4. Asignación del valor de la variable a imputar. 7. Conclusiones. Bibliografía. Anexos.

1. Presentación

El objetivo principal de este trabajo es establecer el proceso metodológico general, operativo y técnico para realizar imputación de valores

faltantes en la variable ingresos. Los resultados que presentamos en este trabajo responden a un primer ejercicio de análisis con el fin de ser discutido con la Dirección General de Estadística y Censos del Gobierno de Ciudad de

¹ El inicio de este artículo se dio en el contexto del proyecto “Metodología para la realización de una Encuesta de Gastos e Ingresos de los hogares de la Ciudad de Buenos Aires” en 2014. Juan Sanguinetti estuvo a cargo de la dirección general del proyecto y el mismo contó con la asistencia técnica de Andrés Meiller. Los autores agradecen los comentarios recibidos a este documento presentado en la sesión del RC55 “*Imputation and Social Indicators: The Use of Factor Analysis for Imputing Missing Data*” del III ISA Forum, realizado en Viena, Austria el 14 de julio de 2016.

Buenos Aires. El ejercicio concreto se aplica a trabajadores por cuenta propia y se utiliza los datos de la Encuesta Nacional de Gastos de los Hogares (ENGHo) 2004-2005 de Argentina.

El trabajo comienza con una breve reseña de los antecedentes y la descripción de los principales procedimientos de imputación, luego se presenta el modelo de análisis, seguido de la metodología empleada y la descripción de los datos utilizados. El apartado de resultados detalla el procedimiento y presenta los hallazgos más importantes. Se concluye con un breve apartado y se complementan el análisis con tres anexos.

2. Antecedentes

Existe una amplia literatura y tradición que aborda la problemática de datos faltantes en donde se han propuesto diferentes tipos de procedimientos de imputación. Los procedimientos varían según el paso del tiempo, desde aquellas decisiones iniciales de eliminación o sustitución de los registros sin información hasta los procedimientos más sofisticados de imputación por simulación.

Los primeros aportes se realizaron en 1932 por Wilks quien propuso el reemplazo de los datos faltantes por la media de los datos presentes de la variable a imputar. Sin embargo, no fue hasta la década del 70 donde comienza el estudio sistemático y la formalización desde un punto de vista probabilístico, destacando principalmente el trabajo de Rubin en 1976 *Inference and missing data*, posteriormente ordenado y sistematizado por Little y Rubin en 1987 en su libro denominado *Statistical Analysis with Missing Data*. Estos autores recogen y sistematizan las evidencias teóricas de imputación de datos otorgando importancia a métodos como los modelos de máxima verosimilitud (Dempster, Laird y Rubin, 1977) y posteriormente a los mecanismos de imputación múltiple (Rubin, 1987).

Por su parte, autores como Droesbeke y Lavallée (1996) hacen un importante aporte a la teoría al analizar y clasificar los distintos tipos de no respuesta. Estos avances en la investigación

permiten estudiar el patrón de comportamiento de los datos omitidos y ello resulta de suma importancia para la elección del modelo de imputación adecuado.

Existe otro conjunto de investigaciones que buscan potenciar los mecanismos de imputación ya desarrollados previamente o bien crear nuevas metodologías como las basadas en redes neuronales (Koikkalainen, 2002), análisis factorial (Genz y Li, 2003; Gleason, y Staelin, 1975) o en la descomposición GH-Biplot (Vásquez, 1995). Todeschini (1990) propuso un k-vecino más cercano como método de estimación de valores perdidos, obteniéndose buenos resultados cuando se cuenta con información auxiliar (Useche y Mesa, 2006).

El trabajo que presentamos aquí se inscribe en esta producción más actual y utiliza como base el análisis factorial de correspondencias múltiples tratando variables categóricas con el objetivo de encontrar el vecino más cercano, según un criterio de distancia como expresión de la proximidad social y criterio de validez. Así, en un espacio multidimensional definido por un conjunto de factores principales derivados de una selección de variables originales inicial, se busca atribuir el valor de la variable faltante con el del individuo que es considerado como el más próximo en este espacio. Se configura así un espacio social en el sentido bourdieano (Bourdieu, 1979). Se trata de un espacio euclidiano y, por un modelo o conjunto de características sociales elegidas y fundamentadas teóricamente, permite valorar quién es el individuo más próximo socialmente y, por tanto, el mejor informante para imputar la información faltante.

3. Procedimientos de imputación

Un elemento importante a tener en cuenta para el proceso de imputación es el reconocimiento de los diversos tipos de distribuciones de los datos faltantes (Rubin, 1977 y Little y Rubin 1987). Estos autores identifican tres tipos de patrones de comportamiento posibles:

- Missing at Random (MAR): cuando la distribución de los valores observados no depende del patrón de comportamiento de los registros sin información.
- Missing not at Random (MNAR): cuando existe dependencia entre los datos completos y los faltantes.
- Missing completely at Random (MCAR): cuando la omisión no depende de los datos observados.

El último comportamiento es el que asumen la mayoría de los algoritmos de imputación.

Useche y Ávila (2006) destacan que los patrones de pérdida de información pueden ser ignorados si ocurren de manera completamente aleatoria (MCAR, Missing Completely at Random) o de manera aleatoria (MAR, Missing At Random). El primer caso (MCAR), ocurre cuando la ausencia de información depende de alguna variable presente en la matriz de datos. El segundo caso (MAR) ocurre cuando la ausencia de los datos depende de variables presentes en la matriz de datos, excluyendo la variable perdida. Los patrones de pérdida no ignorables (NMAR) son los que ocurren cuando la ausencia de los

datos depende de la variable perdida, esto traería como consecuencia estudiar el patrón de pérdida de los datos ausentes para luego imputar tomando en cuenta dicho patrón.

Rivero Rodríguez (2011) en un análisis de la literatura sobre las encuestas sociales registra el hecho de que cuando la variable a imputar es la renta, los datos siguen un patrón MNAR y esto se asocia con el hecho de que los individuos de mayor renta tenderían a ocultarla (Riphahn y Serfling, 2005; Schrapler, 2004).

En un trabajo de imputación de ingresos llevado a cabo para la ENGHo de 1996-97 por Keifman, Manzano, Rodríguez y Viler (1998) comentan que en la práctica es casi imposible determinar si nos encontramos en el caso de no respuesta aleatoria o ignorable, si no se cuenta con información adicional proveniente de otras fuentes tales como re-entrevistas a no respondientes, registros administrativos, etc.

Sin ánimo de ser exhaustivo, la Tabla 1 sintetiza las principales técnicas de imputación y los autores que las crearon junto a los que contribuyeron a su desarrollo.

Tabla 1. Principales técnicas de imputación

Listwise (análisis de datos completos)	•Helmel, 1987
Imputación por medias	•Wilks, 1932
Imputación por medias condicionadas	•Acock y Demo, 1994.
Procedimientos ponderados	•Kalton y Kasprzyk, 1986
Imputación con variables ficticias	•Cohen y Cohen, 1983
Imputación por regresión	•Buck, 1960
Hot-deck	•Fellgi y Holt, 1976 •Mejorado por Madow, Nisselson y Olkin, 1983
Estimación por Máxima Verosimilitud (Algoritmo Expectation-Maximization)	•Dempster, Laird y Rubin, 1977 •McLachlan y Krishnan, 1996; •Little y Rubin, 2002.
Imputación Múltiple (simulación)	•Creado por Rubin en 1987 •Avances: Rubin, 1996; Schafer, 1997; Little y Rubin, 2002; y Zhang, 2003.

Cabe destacar que el tipo de imputación depende también del tipo de encuesta o muestra con la que estemos trabajando. Una serie de autores como Fay (1993), Binder (1996), Binder y Weimin (1996), Montaquila y Jernigan (1997) entre otros, analizan los mecanismos de imputación en encuestas complejas, con diferentes etapas de selección y conglomeración y distintas probabilidades de selección. Dichos autores encuentran que los mecanismos de imputación múltiple (IM) no resultan del todo adecuados para brindar resultados satisfactorios en los estimadores.

Asimismo Medina y Galván (2007) aconsejan actuar con cautela cuando se trabaja con encuestas complejas ya que persiste el desafío de desarrollar algoritmos de imputación robustos que tengan en cuenta el diseño de la muestra y las probabilidades de selección de las observaciones. Estos autores realizan una interesante tipología de los métodos de imputación que resume las características más destacadas de los principales métodos, y señalan los supuestos en que se sustenta cada uno con relación al patrón observado en los datos faltantes, la forma en que se aplica el procedimiento, las principales ventajas y desventajas, y también dan cuenta de la disponibilidad de procedimientos en los paquetes estadísticos SPSS, SAS y STATA. La síntesis de sus hallazgos se presenta en el Anexo 1.

4. Modelo de análisis y metodología

El principio que guía el proceso de imputación es la búsqueda del individuo que más se asemeja socialmente al individuo con falta de información en la variable a imputar, que en el caso que aquí consideramos son los ingresos, variable que se inscribe en una relación de interdependencia con otras variables que definen y acotan un fenómeno o parcela de la realidad social, un espacio social de relaciones. Una vez determinado este individuo más semejante se considera su nivel de ingresos como valor de imputación para el individuo con información faltante. Para determinar esta semejanza se requiere en primer lugar un modelo teórico-empírico que especifique las características

sociales que configuran o estructuran las principales fuentes de variación de la variable a imputar, por tanto, que en mayor medida pueden determinar los distintos niveles de ingresos de los individuos o de los hogares. A continuación, se sigue un procedimiento técnico de búsqueda basado en un criterio de proximidad o “distancia social” donde se considera como medida la distancia euclidiana entre los individuos con información y sin información de la variable a imputar en el espacio euclidiano y social que se obtiene de un análisis factorial de correspondencias múltiples (López-Roldán y Fachelli, 2015, capítulo III.11).

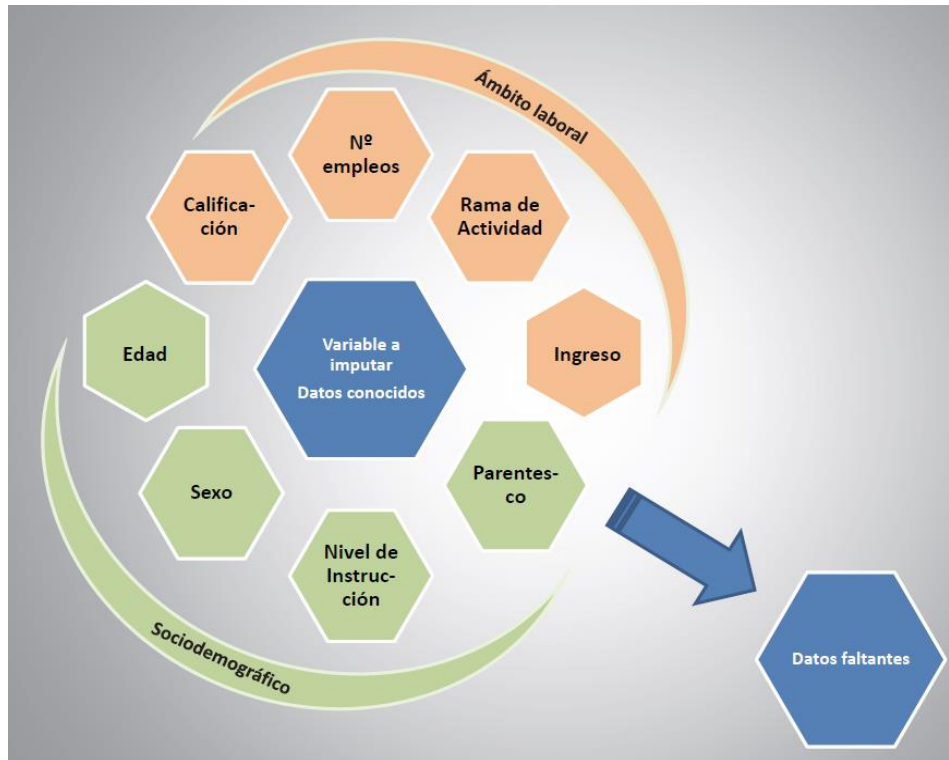
De esta forma se contempla el procedimiento de imputación, que se puede esquematizar en las siguientes etapas:

1. Especificación del modelo teórico-empírico.
2. Procedimiento técnico de imputación.
 - 2.1. Análisis de Correspondencias Múltiples.
 - 2.1.1 Determinación del número de factores e interpretación.
 - 2.1.2 Cálculo de las puntuaciones factoriales.
 - 2.2 Cálculo de la matriz de distancias entre las unidades con y sin datos faltantes.
 - 2.3 Determinación del vecino más cercano a cada unidad con valor faltante.
 - 2.4 Asignación del valor de la variable a imputar.

A continuación, se presenta el modelo de análisis en forma gráfica donde se detallan las variables a utilizar en el ejercicio de imputación propuesto. Cabe destacar que el modelo es una herramienta analítica flexible que es necesario definir en cada caso según las características del dato a imputar y la fuente de datos a utilizar.

En el caso concreto que nos ocupa hemos seleccionado cuatro variables directamente vinculadas al ámbito laboral: calificación, número de empleos, rama de actividad e ingresos. Por otro lado, variables sociodemográficas: edad, sexo, nivel de instrucción y parentesco (Gráfico 1).

Gráfico 1. Especificación del modelo teórico-empírico



Por otro lado, hemos considerado una técnica de análisis factorial como instrumento de tratamiento multivariable de la información seleccionada. Al trabajar con variables categóricas el Análisis Factorial de Correspondencias Múltiples nos permite analizar, describir y representar gráficamente la información de las correspondencias o asociaciones existentes entre el conjunto de variables categóricas originales escogidas y expresarlas de forma combinada y sintética en términos de factores de diferenciación de los individuos.

El Análisis de Correspondencias es una técnica de análisis de datos multivariable que responde a diversas tradiciones y aportaciones. En nuestro caso seguimos los lineamientos de la escuela francesa del análisis de datos (Benzécri, 1973, 1979; Lebart, Morineau y Lambert, 1987; Lebart, Morineau y Piron, 2000; Greenacre, 2008) y la implementación de la técnica en el software SPAD.

En particular el Análisis de Correspondencias Múltiples (ACM) establece las correspondencias

(asociaciones) que existen entre las variables cualitativas. En sentido amplio, se entiende por correspondencia la conexión o relación recíproca entre dos elementos o conjuntos de elementos equivalentes y es una generalización del Análisis Factorial de Correspondencias Simples para el caso de una gran cantidad de variables analizadas simultáneamente (Cornejo, 1988).

Asimismo, el ACM ofrece una formalización geométrica que permite el paso de lo cualitativo a lo cuantitativo, de lo heterogéneo a una construcción de orden estructural (Cornejo, 1988) basado en la comparación de perfiles y en la ordenación espacial de formas, más que de magnitudes (López Roldán, 1994). Seguir los principios del análisis factorial “consiste pues, como en componentes principales, en encontrar los ejes principales de inercia a partir del cálculo de los valores y vectores propios de la nube de puntos mediante la diagonalización de la matriz de varianzas y covarianzas” (López Roldán, 1994).

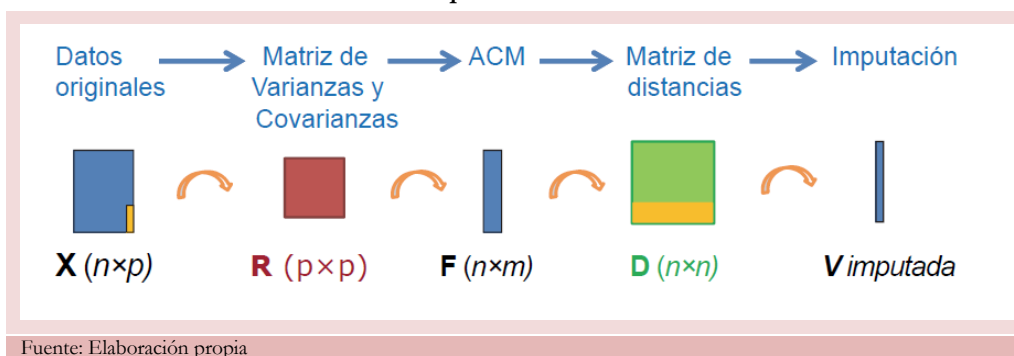
En el ACM intervienen dos tipos de variables, unas denominadas activas que son sobre las cuales opera todo el proceso estadístico para establecer las relaciones de semejanza, y otras ilustrativas, que no intervienen en los cálculos estadísticos pero cuyo nivel de asociación puede ser observado “como si” hubiera participado activamente. De esta manera se puede apreciar gráficamente dónde se habrían ubicado las variables ilustrativas de haber intervenido activamente.

El Gráfico 2 sintetiza el procedimiento técnico llevado adelante. Partiendo de una matriz original de n casos y p variables se construye una matriz de varianzas y covarianzas considerando la distancia de chi-cuadrado para comparar los

perfiles definidos por las distintas categorías correspondientes a las p variables. La aplicación del ACM arroja como resultado los factores o dimensiones principales de inercia, dando cuenta así de la estructura fundamental de las relaciones entre las variables y del fenómeno bajo análisis. A partir de las puntuaciones factoriales se construye una matriz de distancias entre todos los individuos, de los que disponemos de información completa y de los que no la tienen. Así, se eligen los individuos más próximos cuyos ingresos sirven para completar los valores de la variable a imputar.

En el apartado de resultados se detalla este procedimiento en forma aplicada.

Gráfico 2. Procedimiento técnico de imputación



5. Datos

Para dar cuenta del procedimiento realizaremos un análisis de imputación de los ingresos de la población cuentapropista a partir de los datos de la Encuesta Nacional de Gastos de los Hogares de Argentina 2004-2005. Seleccionamos este grupo de trabajadores como ejemplo del proceso de imputación que, por sus características, suelen tener casos perdidos.

Utilizamos específicamente la base de datos de ingresos y delimitamos la población objetivo a cuentapropistas de CABA y Partidos del GBA.

Esta decisión se fundamenta teóricamente en el hecho de que la cercanía geográfica permite el desplazamiento hacia la ciudad de muchos

trabajadores autónomos y empíricamente en el hecho de que la base contaba con 757 cuentapropistas en la ciudad y el análisis con los trabajadores del mismo tipo de GBA permite ampliar la matriz original a 1931 casos.

Para ilustrar el procedimiento consideramos en particular el “Ingreso neto laboral total del preceptor” (variable INGLABT). La tabla siguiente (Tabla 2) presenta los estadísticos de la variable original de la encuesta antes de imputar, donde se pierden 517 casos, y también de la variable con los ingresos imputados por el Instituto Nacional de Estadísticas y Censos (INDEC).

Tabla 2. Estadísticos del Ingreso neto laboral total del preceptor. Submuestra de cuentapropistas

Estadísticos		ENGHo	
		Ingreso neto laboral total del preceptor CON imputados	Ingreso neto laboral total del preceptor SIN imputados
N	Válidos	1,931	1,414
	Perdidos	0	517
Media		912,8	864,1
Mediana		550	500
Moda		0	0
Desv. típ.		1.150,7	1.116,6
Varianza		1.324.074,6	1.246.843,4
Rango		10.655,7	10.655,7
Mínimo		0	0
Máximo		10.655,7	10.655,7
Percentiles	5	20,3	0,6
	10	58,1	50
	15	116,7	100
	20	169,0	150
	25	220	200
	30	283,7	250
	35	350	308,7
	40	400	400
	45	496,0	450
	50	550	500
	55	633,3	600
	60	733,7	700
	65	866,7	800
	70	1.000	966,7
	75	1.183,3	1.067,1
	80	1.400	1.300
	85	1.700	1.600
	90	2.000	2.000
	95	3.000	2.870,8

Fuente: Elaboración propia a partir de la ENGHo 2004-2005

6. Resultados

6.1 Especificación del modelo teórico-empírico

La bondad de la imputación depende sobre todo del conocimiento teórico del fenómeno analizado y de la variable que es objeto de imputación. En el caso de los ingresos el comportamiento de esta variable ha sido analizado en relación a un conjunto de otras variables que se han considerado como relevantes en la determinación de los mayores o menores niveles de ingresos. En concreto las siguientes:

CP26: ¿La semana pasada tenía ... (1 o 2 y más empleos)

CALIFOCUP: Calificación ocupacional de la ocupación principal.

RAMA_PR2: Rama de actividad de la ocupación principal resumida agrupada.

INGLABT220: Ingreso neto laboral total del preceptor sin imputados agrupado en ventiles.

NVINS2: Nivel de Instrucción agrupada.

CP12: Sexo.

EDAD: edad agrupada en intervalos de 10 años.

CP03: Relación de parentesco agrupada.

En consecuencia los resultados de la imputación que se obtengan dependerán en primera instancia del modelo elegido y de las variables empíricas disponibles que lo operativicen. Los resultados que presentamos en este trabajo responden a un primer ejercicio de análisis que requiere su revisión en términos de modelo de análisis y de estudio empírico tomando el caso de la submuestra de cuentapropistas. Como se ha mencionado, el objetivo principal de este trabajo es establecer el proceso metodológico general y el operativo técnico de procesamiento de los datos.

En el Anexo 2 se encuentran, en primer lugar, las frecuencias de las variables de los casos a imputar (517 casos) y, a continuación, la de los casos con ingresos (1.414 trabajadores).

6.2 Procedimiento técnico de imputación

6.2.1 Análisis de correspondencias múltiples

Al operar con esta técnica obtenemos diversos resultados de interés para el objetivo de la imputación de valores faltantes:

- En primer lugar, facilita el tratamiento de variables cualitativas o categóricas.
- El número de variables inicial puede ser numeroso y el número de categorías, con ciertas restricciones, también.
- Como técnica factorial analiza las asociaciones entre el conjunto de variables

inicial y las expresa, de forma combinada y sintética, en factores de diferenciación descomponiendo la inercia total (varianza) en sucesivos factores o ejes que, de manera jerárquica, que van acumulando aquella inercia en términos de factores principales de diferenciación y de estructuración del fenómeno estudiado.

- La lógica del análisis lleva a seleccionar unos pocos factores linealmente independientes que conserven la mayor parte de la inercia total (a partir del 70%), lo que implica una pérdida parcial de información en beneficio de una capacidad reductora y, sobre todo, de la expresión de la variabilidad que es más significativa.
- Al hacerlo pasamos de variables originales que son cualitativas a variables factoriales que son cuantitativas, lo que supone dotar de una métrica a los datos que permite representar a cada categoría de las distintas variables como un punto en el espacio euclidiano.
- De la misma forma que podemos representar los puntos-categorías podemos hacerlo con los puntos-individuos. Es decir, podemos obtener las puntuaciones factoriales de cada individuo en el espacio factorial y por tanto analizar proximidades entre ellos, que interpretamos como distancias sociales definidas por los perfiles que se configuran en los factores.
- En el análisis se pueden considerar variables activas (las que conforman los factores) e ilustrativas o suplementarias que se sitúan en relación al espacio factorial generado. Pero el aspecto más interesante es que podemos realizar el mismo tratamiento con los individuos, distinguiendo los individuos

activos, aquellos en base a los cuales actúan las variables activas seleccionadas, de los individuos ilustrativos, en nuestro caso individuos de los que conocemos una serie de atributos que se expresan en las variables empleadas en el análisis factorial y de los que desconocemos la información de la variable a imputar, en nuestro caso valores perdidos de ingresos. De esta forma podemos posicionar en el espacio factorial a cada individuo del que desconocemos sus ingresos junto al resto de individuos de los que sí conocemos ese dato. La proximidad de puntos en el espacio implica similitud o distancia de individuos según sus perfiles sociales, en relación a todas las variables (factores) consideradas: entre los individuos con información completa en todas las variables y también con respecto a los individuos ilustrativos de los que conocemos toda la información excepto la correspondiente a la variable a imputar.

- En consecuencia, disponemos de un “espacio social” donde los individuos son proyectados y donde sus proximidades se pueden medir para identificar en particular qué individuo con ingresos conocidos es el más similar a otro cuyos ingresos son desconocidos. El ingreso imputado será el ingreso del individuo más parecido o próximo, del vecino más cercano.

Determinación del número de factores e interpretación

Con las 8 variables elegidas se reúne un total de 61 categorías asociadas² que generan un total de 53 dimensiones o factores, una vez eliminadas las categorías con baja frecuencia. De ellos retenemos tan solo dos, puesto que concentran el 70% de la varianza según la transformación de valores propios corregidos propuesta por Benzécri (1979).³

² Inicialmente el número de categorías de las variables era superior. Como algunas de ellas tienen una muy baja frecuencia se requiere revisarlas para su recodificación y agregación con otras categorías. Tras la revisión se consideraron finalmente 65 categorías de las cuales 4 se perdieron en el análisis, pues tienen

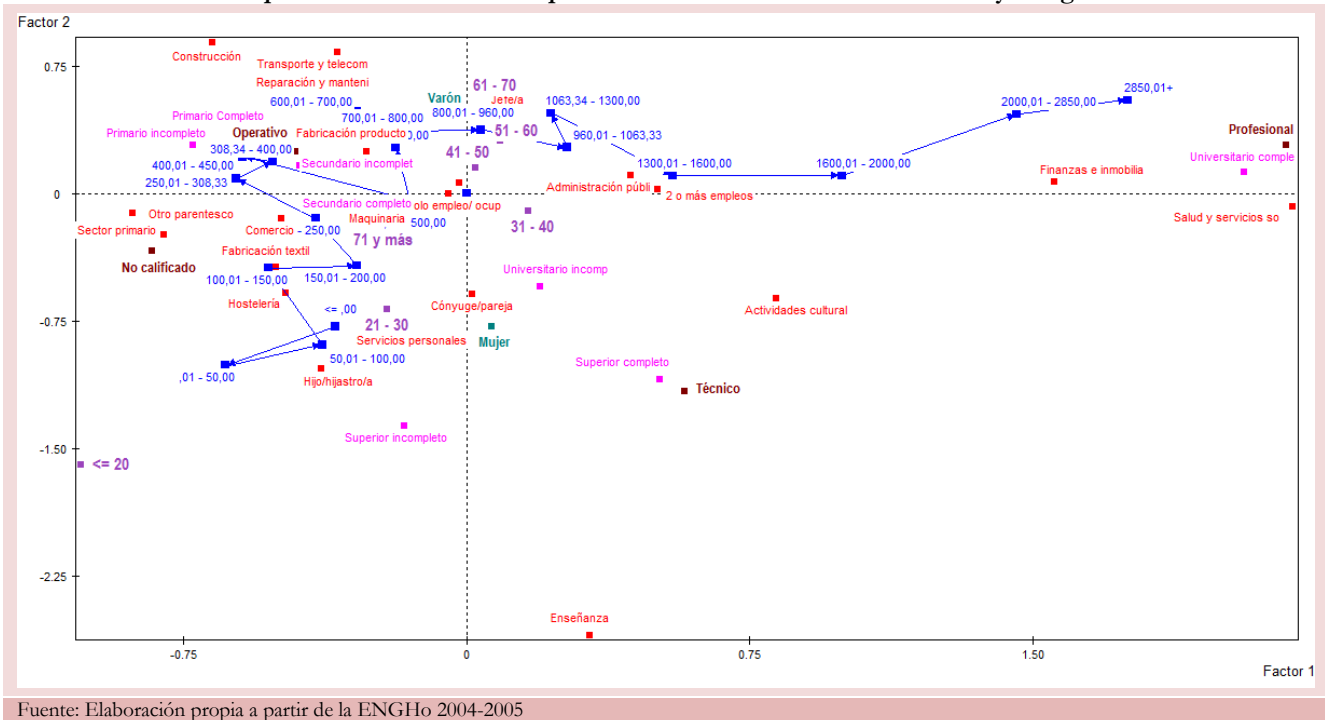
un valor muy poco frecuente, que el propio programa SPAD se encarga de realizar. Estos análisis operativos junto con la revisión del modelo de análisis serán objeto de discusión.

³ A efectos expositivos y para simplificar el análisis se han mantenido estos dos factores en el ejercicio de imputación. No

En el Gráfico 3 se representan las categorías de las variables activas en el espacio factorial de las dos dimensiones retenidas. El primer factor está marcado claramente por la oposición entre niveles de pobreza y riqueza, o de bajos y altos niveles de ingresos junto a los niveles educativos y sectores de actividad más característicos. El segundo factor opone, dentro de los niveles de

ingresos medios-bajos, los perceptores de menores ingresos, donde es mayor la presencia de las mujeres y de determinados sectores de actividad como enseñanza, otros servicios, fabricación, de los perceptores de niveles medios, varones sobre todo, en sectores como construcción, transporte o electricidad.

Gráfico 3. Espacio factorial de las dos primeras dimensiones con las variables y categorías activas



Cálculo de las puntuaciones factoriales

En el espacio factorial obtenido, una categoría es un punto de ese espacio que se ubica en el baricentro de todos los individuos que poseen esa característica. En el Gráfico 4 representamos simultáneamente las categorías de las variables activas (sin etiquetar) en el espacio factorial de las dos dimensiones retenidas con todos los individuos, de acuerdo con sus puntuaciones factoriales. Se representan en azul los individuos activos, cuyos perfiles han configurado los factores, y en color fucsia los individuos ilustrativos de los que no disponemos de información sobre su nivel de ingresos.

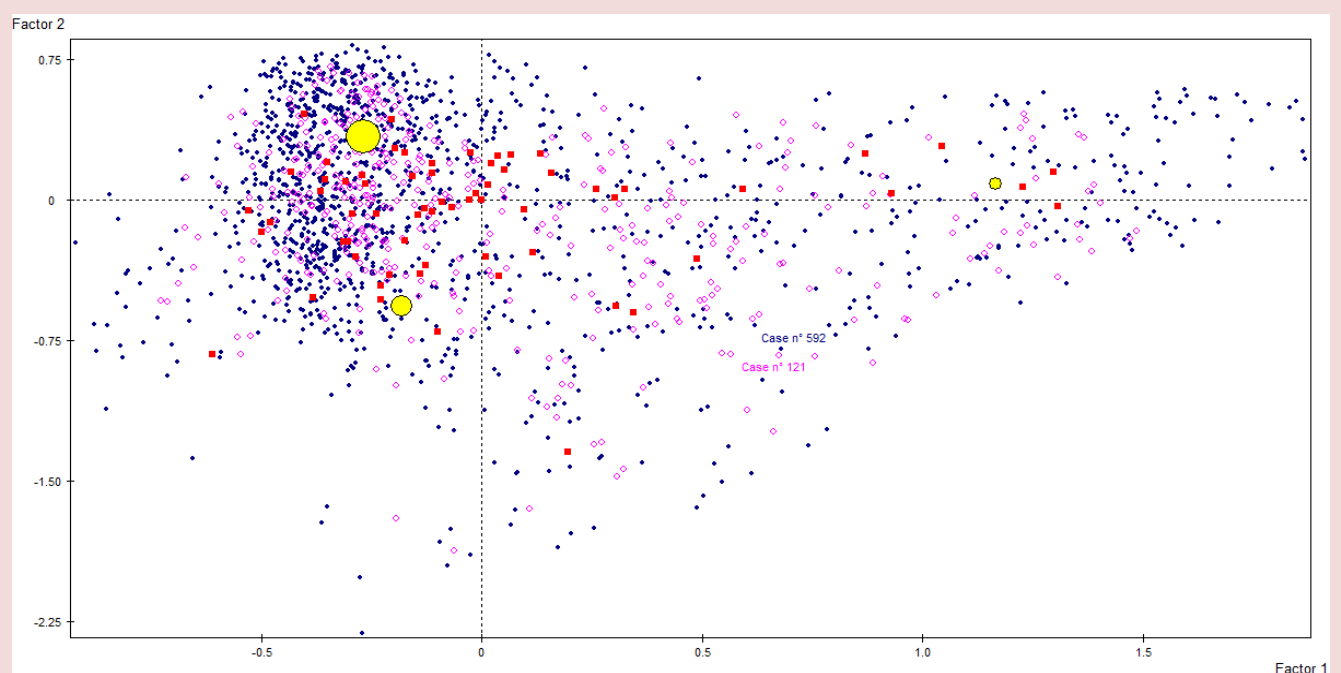
En particular hemos destacado a los individuos 592 (individuo activo) y 121 (individuo ilustrativo). De éste último desconocemos su nivel de ingresos, pero sabemos que se trata de alguien que comparte características sociales asimilables a las del individuo 592, su vecino más cercano. En este caso se trata de dos mujeres de entre 40 y 60 años con estudios superiores completos. De la primera sabemos que ingresa \$1500 de su único trabajo como técnica del sector de la educación. De la segunda sabemos que trabaja como profesional en el sector financiero. La primera no es jefa de hogar y la segunda sí. Estos perfiles se consideran equivalentes imputándose el mismo ingreso de

obstante, podrían incorporarse cualquier número de factores en el proceso de imputación.

\$1500. En el Anexo 3 se presentan los resultados de la aplicación del procedimiento de análisis de

correspondencias múltiples con el software SPAD.

Gráfico 4. Espacio factorial de las dos primeras dimensiones con las variables y categorías activas



Fuente: Elaboración propia a partir de la ENGHo 2004-2005

5.2.2. Cálculo de la matriz de distancias

Para realizar la comparación sistemática entre todos los individuos construimos la matriz de distancias para evaluar sus proximidades. Se trata de buscar el individuo más cercano al individuo a imputar según una medida de distancia. Elegimos la distancia euclidiana.

5.2.3. Determinación del vecino más cercano

En la determinación del vecino más cercano se toma aquel que tiene un valor menor de distancia entre todos los pares comparados, excluyendo de la elección aquellas distancias entre los propios individuos imputados.

5.2.4. Asignación del valor de la variable a imputar

Determinado el individuo con un perfil social más próximo según las variables consideradas en términos factoriales se asigna el valor de ingresos

de éste al individuo que carece del mismo. En la Tabla 2 se compara la variable original de la ENGHo con los datos imputados por el INDEC y la variable imputada por nosotros.

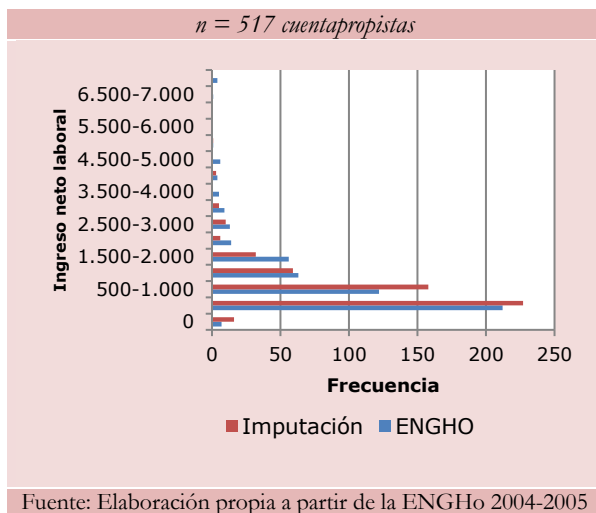
Tabla 2. Distribución agrupada del Ingreso neto laboral total del preceptor. Comparación de la imputación de la ENGHo y la nueva imputación

<i>n = 517 cuentapropistas</i>			
Ingreso	ENGHo	Imputación	Diferencia
0	7	16	9
0-500	212	227	15
500-1.000	122	158	36
1.000-1.500	63	59	-4
1.500-2.000	56	32	-24
2.000-2.500	14	6	-8
2.500-3.000	13	10	-3
3.000-3.500	9	5	-4
3.500-4.000	5	0	-5
4.000-4.500	4	3	-1
4.500-5.000	6	0	-6
5.000-5.500	1	1	0
5.500-6.000	0	0	0
6.000-6.500	0	0	0
6.500-7.000	1	0	-1
+7.000	4	0	-4

Fuente: Elaboración propia a partir de la ENGHo 2004-2005

La diferencia entre los cuentapropistas que han sido imputados en un rango de ingresos por el INDEC y por nuestro ejercicio se presenta en la tercera columna. Cabe destacar que en los extremos de la distribución se observan 16 valores imputados como ingreso cero por el análisis factorial frente a 7 casos imputados por el organismo oficial con otra metodología. Asimismo nuestra imputación finaliza en el rango de ingresos que llega hasta \$5.500, mientras que la del INDEC se extiende hasta más de \$7.000. El gráfico comparativo de ambas imputaciones se presenta a continuación (Gráfico 5).

Gráfico 5. Distribución agrupada del Ingreso neto laboral total del preceptor. Comparación de la imputación de la ENGHo y la nueva imputación



La Tabla 3 presenta las diferencias en la media entre ambas imputaciones, mientras la media del INDEC arroja un valor de \$1.046 nuestro análisis imputa \$750.

La varianza presenta un valor más reducido en el caso de nuestra imputación, así como también el rango de la variable que es de \$5.033 frente a \$9.950 que es el definido por el Instituto.

Tabla 3. Estadísticos del Ingreso neto laboral total del preceptor. Comparación de la imputación de la ENGHo y la nueva imputación

<i>n = 517 cuentapropistas</i>				
	ENGHO	Nueva Imputación	Diferencia	
N	Válidos	517	517	0
	Perdidos	0	0	0
Media	1.046,0	750,5		295,4
Mediana	663,3	540,0		123,3
Moda	500	500		0
Desv. típ.	1.230,4	720,2		510,2
Varianza	1.513.846,0	518.637,6		995.208,4
Rango	9.950,0	5.033,3		4.916,7
Mínimo	0	0		0
Máximo	9.950,0	5.033,3		4.916,7
Percentiles	5	50,0	40,0	10,0
	10	105,3	100,0	5,3
	15	164,2	160,0	4,2
	20	233,3	200,0	33,3
	25	294,2	250,0	44,2
	30	350,0	358,3	-8,3
	35	400,0	403,3	-3,3
	40	500,0	500,0	0,0
	45	584,0	500,0	84,0
	50	663,3	540,0	123,3
	55	796,7	600,0	196,7
	60	900,0	666,7	233,3
	65	1.000,0	717,5	282,5
	70	1.183,3	850,0	333,3
	75	1.400,0	1.000,0	400,0
	80	1.628,3	1.133,3	495,0
	85	1.868,2	1.360,0	508,2
	90	2.426,0	1.603,3	822,7
	95	3.340,1	2.023,3	1.316,8

Fuente: Elaboración propia a partir de la ENGHo 2004-2005

Como se puede observar en la Tabla 4, que incluye todos los casos de la submuestra de cuentapropistas, al incluir la nueva imputación, se observan valores inferiores de ingresos, haciendo que la media pase de \$912,8 a \$833,7.

En términos de dispersión, la nueva imputación presenta una desviación típica menor (de \$1.150,7 a \$1.026,7) si bien el coeficiente de variación experimenta un leve aumento de 0,79 a 0,81. Estas diferencias vienen dadas por el comportamiento diferenciado del resto de la provincia de Buenos Aires, pues la Ciudad Autónoma de Buenos Aires mantiene prácticamente los mismos valores en el coeficiente de variación entre las dos formas de imputación: 0,855 y 0,851.

Tabla 4. Estadísticos del Ingreso neto laboral total del preceptor según territorio. Comparación de la imputación de la ENGHo y la nueva imputación. Submuestra de cuentapropistas

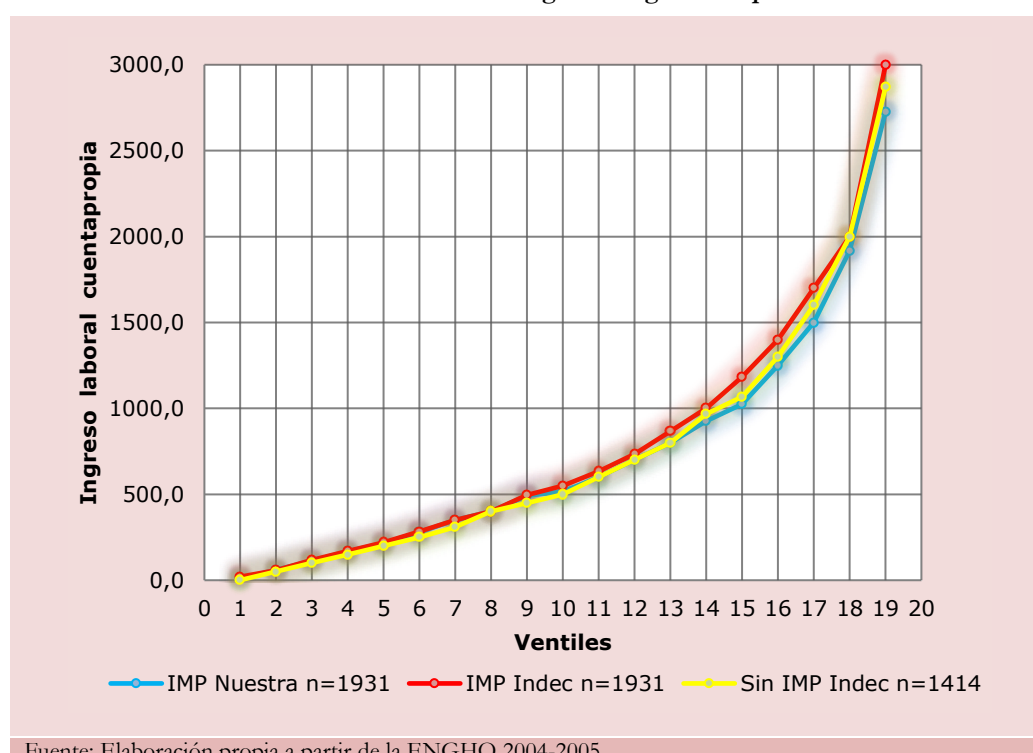
Estadísticos	Ciudad de Buenos Aires		Buenos Aires		Total	
	ENGHo	Nueva Imputación	ENGHo	Nueva Imputación	ENGHo	Nueva Imputación
N	757,00	757,00	1.174,00	1.174,00	1.931,00	1.931,00
Media	1.240,10	1.141,05	701,69	635,45	912,76	833,66
Mediana	833,33	733,33	451,67	460,83	550,00	516,67
Desv. típ.	1.450,43	1.341,36	842,00	688,73	1.150,68	1.026,66
Varianza	2.103.733,27	1.799.236,52	708.971,65	474.348,72	1.324.074,62	1.054.033,27
Mínimo	0,00	0,00	0,00	0,00	0,00	0,00
CV	0,85	0,85	0,83	0,92	0,79	0,81
Máximo	10.655,67	10.655,67	9.375,00	7.000,00	10.655,67	10.655,67
Rango	10.655,67	10.655,67	9.375,00	7.000,00	10.655,67	10.655,67

Fuente: Elaboración propia a partir de la ENGHo 2004-2005

Finalmente, se presenta a continuación la distribución de los cuentapropistas en función de las dos imputaciones (INDEC y la propia) más los casos que declaran ingresos con el fin de compararlas.

Aunque los ventiles iniciales no presentan diferencias, éstas aumentan entre el ventíl 13 y 18. De todas maneras se puede advertir que no hay grandes variaciones.

Gráfico 6. Curva de Lorenz. Distribución de ingresos según la imputación



Fuente: Elaboración propia a partir de la ENGHo 2004-2005

7. Conclusiones

El ejercicio presentado en este informe se erige como una propuesta de imputación a discutir con el objetivo de sentar las bases de un nuevo mecanismo de imputación que complemente las metodologías existentes a partir de ampliar el abanico de posibilidades explorando metodologías elaboradas sobre la base del análisis factorial antes que las metodologías más tradicionales mayormente fundamentadas en funciones de regresión.

Del conjunto de métodos de imputación anteriormente comentado y recopilado por Medina y Galván (2007) seleccionamos el *hotdeck* para realizar un cuadro comparativo en función del método diseñado por nosotros (Tabla 5).

Esta metodología tiene el atractivo de materializar las nociones de “espacio social” y de “configuración” que remiten a una conjunción de factores, es decir, a la aparición conjunta de distintas dimensiones de lo social que se refuerzan y resignifican al ser analizadas de forma integrada. Se trata de una metodología no determinista que busca comprender cómo se relacionan distintos factores a partir de observar la forma en que interactúan las distintas variables (Fachelli et al., 2012).

Así el ámbito particular que conforma el mercado de trabajo donde se desenvuelve el cuentapropista como colectivo particular exhibe en forma entrelazada las distintas dimensiones de lo social y lo laboral pues aborda el conjunto de atributos en el que se expresan y diferencian los perfiles de trabajadores en apariencia similares (trabajadores por cuenta propia).

Tabla 5. Comparación de métodos de imputación

Método	Hotdeck con regresión (condicionado a covariables)	Metodología de imputación según el criterio de proximidad en un espacio factorial
Supuestos y patrón de datos faltante	Patrón de datos faltantes MCAR. Se requiere especificar un modelo, en donde las covariables estén altamente correlacionadas con la variable a imputar.	Patrón de datos faltantes MCAR. Se especifica un modelo factorial, donde las variables están correlacionadas con la variable a imputar.
Aplicación	Se divide la base de datos en subgrupos utilizando variables correlacionadas. Los valores faltantes se sustituyen con el valor medio estimado por la regresión efectuada en el subgrupo de interés.	Se realiza el tratamiento conjunto de los datos a imputar del grupo social considerado. Los valores faltantes se sustituyen por un donante que es el vecino más cercano en el espacio factorial de m dimensiones.
Ventajas	Fácil de aplicar.	La modelización teórica es muy flexible. Posibilidad de aplicar a variables cualitativas. Se pueden utilizar tantas variables como se desee.
Desventajas	Todas las observaciones pertenecientes al subgrupo tienen el mismo valor imputado. Se subestima la varianza y se introducen sesgos en la correlación. No siempre es fácil encontrar un modelo adecuado para la variable de interés. Se subestima el error estándar del estimador.	Obtención de diferentes resultados en función del modelo seleccionado, el número de factores retenidos y de la medida de distancia elegida.
Software	STATA	SPSS y SPAD

En ese espacio de diferenciación sociolaboral hay casos muy similares, que comparten coordenadas factoriales, y por ende se constituyen en vecinos donantes de un atributo concreto como es el ingreso a otro individuo que compartiendo el mismo “espacio social” no declaró los suyos. No obstante, la potencialidad de la herramienta metodológica empleada en este ejercicio permite que sea utilizada para imputar valores faltantes de cualquier tipo de variable, trascendiendo por tanto la aplicación exclusiva al ingreso.

Como se comentó anteriormente, esta propuesta se pone a discusión, considerando que algunos temas pendientes son los siguientes:

- Definición del Modelo de Análisis en función de la experiencia de la Dirección en el tratamiento de los ingresos.
- Definición de los diferentes niveles de análisis para la imputación: por categorías ocupacionales, por ámbitos geográficos, etc.
- Exploración y generación de indicadores de confianza sobre los procedimientos de imputación empleados.

Finalmente, una línea de trabajo interesante a desarrollar podría ser un análisis sistemático de aplicación de los diferentes ejercicios de imputación (modelos clásicos) y su comparación con el que se presenta en el cuerpo de este trabajo. Sería una de las maneras posibles, entre otras, de difundir el interés y compromiso que ha manifestado la institución en la mejora de las herramientas metodológicas para el tratamiento de los datos de encuesta.

Bibliografía

- Acock, C. A., y Demo, D. (1994). *Family diversity and well-being*. Thousand Oaks, C. A. Sage.
- Benzécri, J. P. (1973). *L'Analyse des données. II. L'analyse des correspondances*. Paris: Dunod.

- Benzécri, J. P. (1979). Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Les Cahiers De l'Analyse Des Données*, 3, IV, 377-388.
- Binder, D. A. (1996). Comment to articles by Rao, Fay and Rubin. *Journal of the American Statistical Association* 91.
- Binder, D. A. y Sun, W. (1996). *Frequency valid multiple imputation for surveys with complex designs*. Business Survey Methods Division, Statistics, Canada.
- Bourdieu, P. (1979). *La distinction. Critique sociale du jugement*. Paris: Les Editions de Minuit.
- Buck, S. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer, *Journal of the Royal Statistical Society*, 22, 2, 302-307.
- Cohen, J., y Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, N. J., Erlbaum.
- Cornejo, J. M. (1988): *Técnicas de Investigación Social: El análisis de correspondencias (Teoría y Práctica)*. Barcelona: Promociones y Publicaciones Universitarias.
- Dempster, A. P., Lair, N. M. y Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39.
- Droesbeke, J. J. y Lavallée, P. (1996). La no respuesta en las encuestas. Metodológica N° 4.
- Fachelli, S., López, N; López-rolán, P. y Sourrouille, F. (2012) *Desigualdad y diversidad en América Latina: Hacia un análisis tipológico comparado*. Buenos Aires: IIPE-UNESCO.
- Fay, R. E. (1993). *Valid Inference from imputed survey data*. Washington, D. C.: U. S. Bureau of the Census.
- Fellegi, I. y Holt, D. (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*. Vol. 71, 353, 17-35.
- Geng, Z. y Li, K. (2003). Factorization of posteriors and partial imputation algorithm for graphical models with missing data. *Statistics and probability letters*, 64, 369-379.
- Gleason, T. y Staelin, R. (1975). *A proposal for handling missing data*. *Psychometrika*. Vol. 40, 2, 229-252.

- Greenacre, Michael (2008) *La práctica del análisis de correspondencias*. Manuales Fundación BBVA. Madrid.
- Goicoechea, P. (2002). Imputación basada en árboles de clasificación. EUSTAT.
- Government Statistical Service. (1996). Report of the Task Force on Imputation. *GSS Methodology Series*, 3.
- Hemel, J. y Otros (1987). Stepwise deletion: a technique for missing data handling in multivariate analysis. *Analytical Chemical Acta* 193, 255-268.
- Kalton, G. y Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology*, 12.
- Keifman, S.; Manzano, G.; Rodríguez, C. y Vilker, A. (1998) Imputación de Ingresos de Hogares: La Experiencia de la Encuesta Nacional de Gastos de los Hogares de Argentina. *Informe del 2º Taller Regional MECOVI. INDEC/CEPAL*, Buenos Aires.
- Koikkalainen, P. (2002). *Neural Network for editing and imputation*. University of Jyväskylä Finland.
- Lebart, L.; Morineau, A. y Piron, M. (2000). *Statistique exploratoire multidimensionnelle*. Dunod: Paris.
- Lebart L. ; Morineau A. y Lambert, T. (1987) *SPAD.N: Système Portable pour l'Analyse des Données. Version 1.0. Manuel de Référence*. Sèvres, France: CISIA.
- Little, R. J., y Rubin, D. (1987). *Statistical analysis with missing data*. New York, Wiley.
- Little, R. J., y Rubin, D. (2002). *Statistical analysis with missing data (Sec. Ed.)*. New York, Wiley.
- López-Roldán, P. (1996): La construcción de tipologías: metodología de análisis, *Revista Papers*, 48, 9-29.
- López-Roldán, P. (1994): *La construcción de tipologías en sociología: propuesta metodológica de construcción, análisis y validación. Aplicación al estudio de la segmentación del mercado de trabajo en la Región Metropolitana de Barcelona*. Tesis Doctoral. Universitat Autònoma de Barcelona. Bellaterra (en línea) <http://www.tesisexarxa.net/TDX-0507108-161640/#documents>
- López-Roldán, P.; Fachelli, S. (2015). *Metodología de la Investigación Social Cuantitativa*. Bellaterra (Cerdanyola del Vallès): Dipòsit Digital de Documents, Universitat Autònoma de Barcelona. 1ª edición. Edición digital: <http://ddd.uab.cat/record/129382>.
- Madow, W. G., Nisselson, J. y Olkin, I. (Eds.) (1983). *Incomplete data in sample surveys*. Vol. 1, Report and case studies. New York: Academic Press.
- McLachlan, G.J. y Krishnan, T., (1996). *The EM algorithm and extensions*. New York: Wiley.
- Medina, F. y Galván, M. (2007). Imputación de datos: teoría y práctica. *CEPAL, Serie de estudios estadísticos y prospectivas*, 54.
- Montaquila, M. J. y R. Jernigan W. (1997). Variance estimation in the presence of imputed data. *Proceedings of the Survey Research Methods Section of the American Statistical Association*.
- Platek, R. (1986). Metodología y Tratamiento de la no respuesta. *Seminario Internacional de Estadística en EUSKADI*. Cuaderno 10.
- Rivero Rodríguez, G. (2011) *Análisis de datos incompletos en Ciencias Sociales*. Madrid: CIS. Cuadernos Metodológicos, 046.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* 63.
- Rubin, D. B. (1977). Formalizing subjective notions about the effect of non-respondents in sample surveys. *Journal of the American Statistical Association* 72
- Rubin, D. B. (1987). *Multiple imputation for non-response in surveys*. New York, Wiley.
- Rubin, D. B. (1996). Multiple Imputation after 18+ years (with discussion). *Journal of the American Statistical Association* 91.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London, Chapman y Hall.
- Useche Castro, L. M. y Mesa Ávila, M. (2006). Una introducción a la Imputación de Valores Perdidos. *Terra Nueva Etapa*, vol. XXII, 31, 2006, 127-152.

Anexo 1: Tipología de los métodos para la imputación de datos

Método	Supuestos y patrón de datos faltante	Aplicación	Ventajas	Desventajas	SPSS	SAS	STATA
Determinístico	Patrón de datos faltantes condicionado (MNAR)	Los datos faltantes se imputan conforme a las reglas definidas en los catálogos de crítica y codificación.	Los datos se imputan con la información del registro que se analiza, lo cual garantiza consistencia.	Las situaciones que ocurren en la realidad superan la imaginación del personal encargado de preparar criterios ad-hoc. No es posible generar criterios determinísticos para todas las variables.	*	*	*
Repetición	Patrón de datos faltantes condicionado (MNAR)	En las encuestas de hogares continuas se utiliza información pasada de la misma unidad de observación.	Se hace uso de información de las mismas unidades de observación.				
Datos completos (Listwise)	Patrón de datos faltantes MCAR o MAR. Los datos eliminados son una submuestra aleatoria de la muestra.	Se utilizan sólo registros con información.	Fácil de aplicar. Es la opción que por defecto aplican los paquetes SPSS, SAS, STATA	Se reduce el tamaño de la muestra. Se debilita la significancia estadística de la pruebas.	X	X	X
Datos disponibles (Pairwise)	Patrón de datos faltantes MCAR.	Se utilizan las observaciones que tienen información para todas las variables.	Fácil de aplicar y trabaja con toda la información.	Se distorsiona la relación entre variables.	X		
Medias no condicionadas	Patrón de datos faltantes MCAR.	Las observaciones faltantes se reemplazan por el valor medio de la variable de análisis.	Fácil de entender y aplicar.	Genera estimadores sesgados. Subestima la varianza de los estimadores.	X	X	
Medias por subgrupos	Patrón de datos faltantes MCAR.	Se divide la base de datos en subgrupos utilizando variables correlacionadas. Las observaciones faltantes en el subgrupo de interés se reemplazan por el valor medio de la variable.		Genera estimadores sesgados. Subestima la varianza de los estimadores.	X	X	

Variables binarias	Patrón de datos faltantes MCAR.	Se forman variables binarias para identificar las observaciones con datos faltantes.		Introducen distorsiones en la interpretación de los parámetros de los modelos de regresión.	X		
Reponderación	Patrón de datos faltantes MCAR.	Se forman grupos, y al interior de los mismos se eliminan las observaciones sin datos. Se ponderan las observaciones utilizando la información de la muestra completa. Los factores de reponderación se pueden obtener por modelos.	Las observaciones que permanecen en muestra y los ponderadores ajustados estiman correctamente los totales por celda.	Cuando la falta de respuesta por celda es muy alta, es posible que se introduzcan sesgos en el valor de los estimadores y su varianza			
Hotdeck (condicionado a covariables)	Patrón de datos faltantes MAR.	Se divide la base de datos en subgrupos utilizando variables correlacionadas. Los valores faltantes se sustituyen con la información de un registro con información similar en las covariables.	Los donantes y receptores pertenecen a un mismo subgrupo. El número de donantes se condiciona con el uso de covariables.	No siempre es fácil definir un criterio de distancia o similitud entre los posibles donantes y receptores.			X
Hotdeck con regresión (condicionado a covariables)	Patrón de datos faltantes MCAR. Se requiere especificar un modelo, en donde las covariables estén altamente correlacionadas con la variable a imputar.	Se divide la base de datos en subgrupos utilizando variables correlacionadas. Los valores faltantes se sustituyen con el valor medio estimado por la regresión efectuada en el subgrupo de interés.	Fácil de aplicar.	Todas las observaciones pertenecientes al subgrupo tienen el mismo valor imputado. Se subestima la varianza y se introducen sesgos en la correlación. No siempre es fácil encontrar un modelo adecuado para la variable de interés. Se subestima el error estándar del estimador.			X

Regresión	Patrón de datos faltantes MCAR. Se requiere especificar un modelo, en donde las covariables estén altamente correlacionadas con la variable a imputar.			Se subestima el error estándar del estimador. Se subestima la varianza.	X		X
Regresión por subgrupos	Patrón de datos faltantes MCAR. Se requiere especificar un modelo, en donde las covariables estén altamente correlacionadas con la variable a imputar.	Se divide la base de datos en subgrupos utilizando variables correlacionadas. Los valores faltantes se sustituyen con el valor medio estimado por la regresión efectuada en el subgrupo de interés.	Fácil de aplicar.	Genera estimadores sesgados. Subestima la varianza de los estimadores. Genera sesgos de correlación. No siempre es fácil encontrar un modelo adecuado para la variable de interés. Se subestima el error estándar del estimador.	X		X
Regresión aleatoria	Patrón de datos faltantes MAR. Se requiere especificar un modelo en donde las covariables estén altamente correlacionadas con la variable a imputar.			No siempre es fácil encontrar un modelo adecuado para la variable de interés. Se subestima el error estándar del estimador. Se subestima la varianza.			X
Máxima Verosimilitud (EM)	Patrón de datos faltantes MAR.		Genera estimaciones robustas basadas en la muestra observada. No efectúa simulaciones.	No siempre están disponibles. Hay que programar el algoritmo que se desea aplicar.	X	X	

Imputación simple	Patrón de datos faltantes MAR. Se requiere especificar un modelo en donde las covariables estén altamente correlacionadas con la variable a imputar.		Utiliza procedimientos estadísticamente robustos.	Requiere que los datos faltantes sigan el patrón MAR. Ocurren situaciones en que los supuestos del método no se cumplen. No es posible conocer el error estándar de los estimadores, ya que sólo efectúa un interacción. No siempre es fácil encontrar un modelo adecuado para la variable de interés.			X
Imputación múltiple	Patrón de datos faltantes MAR. Se requiere especificar un modelo, en donde las covariables estén altamente correlacionadas con la variable a imputar. El modelo que se utiliza para imputar, debe ser el mismo o muy similar al que se utilizará en el análisis secundario de datos.	Se generan m subconjuntos de datos imputados por medio de simulaciones. Se combina en forma apropiada a fin de obtener estimadores robustos.	Muy intuitivo y fácil de entender. Utiliza procedimientos estadísticos robustos. Genera distintas opciones de datos imputados y las combina en forma adecuada. Permite calcular el error estándar de los estimadores.	Requiere que los datos faltantes sigan el patrón MAR. Ocurren situaciones en que los supuestos del método no se cumplen. No siempre es fácil encontrar un modelo adecuado para la variable de interés. Se requiere de paquetes estadísticos de cómputo que contengan algoritmos de cálculo para este propósito. Si no se utilizan con precaución pueden funcionar como cajas negras. Se le deja la responsabilidad a un procedimiento estadístico.		X	X

Fuente: Medina, F. y Galván, M. (2007)

* Programando

Anexo 2: Frecuencias de las variables seleccionadas para la elaboración del modelo.

Estadísticos Cuentalpropistas a imputar

CP26 ¿La semana pasada tenía ...					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 un solo empleo/ ocupación?	483	93,4	93,4	93,4
	2 2 o más empleos	34	6,6	6,6	100,0
	Total	517	100,0	100,0	

CALIFOCUP Calificación ocupacional de la ocupación principal					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Profesional	67	13,0	13,0	13,0
	2 Técnico	84	16,2	16,2	29,2
	3 Operativo	344	66,5	66,5	95,7
	4 No calificado	19	3,7	3,7	99,4
	5 Ns/Nc	3	,6	,6	100,0
	Total	517	100,0	100,0	

RAMA_PR2 Rama de actividad de la ocupación principal resumida					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00 Sector primario	15	2,9	2,9	2,9
	2,00 Fabricación productos	19	3,7	3,7	6,6
	3,00 Fabricación textil	15	2,9	2,9	9,5
	4,00 Maquinaria	19	3,7	3,7	13,2
	5,00 Construcción	65	12,6	12,6	25,7
	6,00 Comercio	136	26,3	26,3	52,0
	7,00 Reparación y mantenimiento	28	5,4	5,4	57,4
	8,00 Hostelería	16	3,1	3,1	60,5
	9,00 Transporte y telecomunicaciones	40	7,7	7,7	68,3
	10,00 Finanzas e inmobiliarias	85	16,4	16,4	84,7
	11,00 Administración pública	1	,2	,2	84,9
	12,00 Enseñanza	19	3,7	3,7	88,6
	13,00 Salud y servicios sociales	16	3,1	3,1	91,7
	14,00 Actividades culturales	21	4,1	4,1	95,7
	15,00 Servicios personales	21	4,1	4,1	99,8
	16,00 Ns/Nc	1	,2	,2	100,0
	Total	517	100,0	100,0	

NIVINS2 Nivel de instrucción					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00 Primario incompleto	41	7,9	7,9	7,9
	2,00 Primario Completo	119	23,0	23,0	30,9
	3,00 Secundario incompleto	91	17,6	17,6	48,5
	4,00 Secundario completo	90	17,4	17,4	66,0
	5,00 Superior incompleto	14	2,7	2,7	68,7
	6,00 Superior completo	37	7,2	7,2	75,8
	7,00 Universitario incompleto	51	9,9	9,9	85,7
	8,00 Universitario completo	74	14,3	14,3	100,0
	Total	517	100,0	100,0	

CP12 Sexo					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Varón	344	66,5	66,5	66,5
	2 Mujer	173	33,5	33,5	100,0
	Total	517	100,0	100,0	

Edad_agrupada Edad (agrupado)					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 <= 20	16	3,1	3,1	3,1
	2 21 - 30	88	17,0	17,0	20,1
	3 31 - 40	98	19,0	19,0	39,1
	4 41 - 50	134	25,9	25,9	65,0
	5 51 - 60	102	19,7	19,7	84,7
	6 61 - 70	52	10,1	10,1	94,8
	7 71 y más	27	5,2	5,2	100,0
	Total	517	100,0	100,0	

CP03 Relación de parentesco					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Jefe/a	297	57,4	57,4	57,4
	2 Cónyuge/pareja	101	19,5	19,5	77,0
	3 Hijo/hijastro/a	93	18,0	18,0	95,0
	4 Otro parentesco	26	5,0	5,0	100,0
	Total	517	100,0	100,0	

Estadísticos Cuentapropistas con ingresos

CP26 ¿La semana pasada tenía ...					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 un solo empleo/ ocupación?	1293	91,4	91,4	91,4
	2 2 o más empleos	121	8,6	8,6	100,0
	Total	1414	100,0	100,0	

CALIFOCUP Calificación ocupacional de la ocupación principal					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Profesional	167	11,8	11,8	11,8
	2 Técnico	224	15,8	15,8	27,7
	3 Operativo	942	66,6	66,6	94,3
	4 No calificado	74	5,2	5,2	99,5
	5 Ns/Nc	7	,5	,5	100,0
	Total	1414	100,0	100,0	

RAMA_PR2 Rama de actividad de la ocupación principal resumida					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00 Sector primario	28	2,0	2,0	2,0
	2,00 Fabricación productos	54	3,8	3,8	5,8
	3,00 Fabricación textil	78	5,5	5,5	11,3
	4,00 Maquinaria	69	4,9	4,9	16,2
	5,00 Construcción	193	13,6	13,6	29,8
	6,00 Comercio	343	24,3	24,3	54,1
	7,00 Reparación y mantenimiento	85	6,0	6,0	60,1
	8,00 Hostelería	27	1,9	1,9	62,0
	9,00 Transporte y telecomunicaciones	113	8,0	8,0	70,0
	10,00 Finanzas e inmobiliarias	192	13,6	13,6	83,6
	11,00 Administración pública	4	,3	,3	83,9
	12,00 Enseñanza	57	4,0	4,0	87,9
	13,00 Salud y servicios sociales	65	4,6	4,6	92,5
	14,00 Actividades culturales	34	2,4	2,4	94,9
	15,00 Servicios personales	71	5,0	5,0	99,9
	16,00 Ns/Nc	1	,1	,1	100,0
	Total	1414	100,0	100,0	

NIVINS2 Nivel de instrucción					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00 Primario incompleto	144	10,2	10,2	10,2
	2,00 Primario Completo	336	23,8	23,8	33,9
	3,00 Secundario incompleto	242	17,1	17,1	51,1
	4,00 Secundario completo	260	18,4	18,4	69,4
	5,00 Superior incompleto	36	2,5	2,5	72,0
	6,00 Superior completo	83	5,9	5,9	77,9
	7,00 Universitario incompleto	119	8,4	8,4	86,3
	8,00 Universitario completo	194	13,7	13,7	100,0
	Total	1414	100,0	100,0	

CP12 Sexo					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Varón	876	62,0	62,0	62,0
	2 Mujer	538	38,0	38,0	100,0
	Total	1414	100,0	100,0	

Edad_agrupada Edad (agrupado)					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 <= 20	39	2,8	2,8	2,8
	2 21 - 30	190	13,4	13,4	16,2
	3 31 - 40	293	20,7	20,7	36,9
	4 41 - 50	332	23,5	23,5	60,4
	5 51 - 60	335	23,7	23,7	84,1
	6 61 - 70	164	11,6	11,6	95,7
	7 71 y más	61	4,3	4,3	100,0
	Total	1414	100,0	100,0	

CP03 Relación de parentesco					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1 Jefe/a	851	60,2	60,2	60,2
	2 Cónyuge/pareja	348	24,6	24,6	84,8
	3 Hijo/hijastro/a	175	12,4	12,4	97,2
	4 Otro parentesco	40	2,8	2,8	100,0
	Total	1414	100,0	100,0	

Anexo 3: Resultados del Análisis de Correspondencias Múltiples

SELECTION OF CASES AND VARIABLES

ACTIVE CATEGORICAL VARIABLES

8 VARIABLES 65 ASSOCIATED CATEGORIES

8 . ¿La semana pasada tenía ...	(2 CATEGORIES)
32 . Relación de parentesco	(4 CATEGORIES)
33 . Sexo	(2 CATEGORIES)
39 . Calificación ocupacional de la ocupación principal	(5 CATEGORIES)
47 . Edad (agrupado)	(7 CATEGORIES)
51 . Ingreso neto laboral total del preceptor sin imputados (agru	(21 CATEGORIES)
52 . Rama de actividad de la ocupación principal resumida	(16 CATEGORIES)
53 . Nivel de instrucción	(8 CATEGORIES)

SUPPLEMENTARY CATEGORICAL VARIABLES

1 VARIABLES 3 ASSOCIATED CATEGORIES

17 . Marca del ingreso laboral total del preceptor	(3 CATEGORIES)
--	------------------

SUPPLEMENTARY CONTINUOUS VARIABLES

1 VARIABLES

16 . Ingreso neto laboral total del preceptor	(CONTINUOUS)
---	----------------

CASES

WEIGHT OF CASES	: Weight of objects, uniform equal to 1.	UNIF
KEPT	NITOT = 1931 PITOT = 1931.000	
SELECTION AFTER FILTERING		
ACTIVE	NIACT = 1417 PIACT = 1417.000	
SUPPLEMENTARY	NISUP = 514 PISUP = 514.000	

MULTIPLE CORRESPONDENCE ANALYSIS

ELIMINATION OF ACTIVE CATEGORIES WITH SMALL WEIGHTS

THRESHOLD (PCMIN) : 1.80 % WEIGHT: 25.51
 BEFORE CLEANING : 8 ACTIVE QUESTIONS 65 ASSOCIATE CATEGORIES
 AFTER CLEANING : 8 ACTIVE QUESTIONS 61 ASSOCIATE CATEGORIES
 TOTAL WEIGHT OF ACTIVE CASES : 1417.00
 MARGINAL DISTRIBUTIONS OF ACTIVE QUESTIONS

IDENT	CATEGORIES LABEL	BEFORE CLEANING COUNT	BEFORE CLEANING WEIGHT	AFTER CLEANING COUNT	AFTER CLEANING WEIGHT	HISTOGRAM OF RELATIVE WEIGHTS,
8 . ¿La semana pasada tenía ...						
CP01 - un solo empleo/ ocup		1296	1296.00	1296	1296.00	
CP02 - 2 o más empleos		121	121.00	121	121.00	*****
32 . Relación de parentesco						
CP01 - Jefe/a		854	854.00	854	854.00	*****
CP02 - Cónyuge/pareja		348	348.00	348	348.00	*****
CP03 - Hijo/hijastro/a		175	175.00	175	175.00	*****
CP04 - Otro parentesco		40	40.00	40	40.00	**
33 . Sexo						
CP01 - Varón		876	876.00	876	876.00	*****
CP02 - Mujer		541	541.00	541	541.00	*****
39 . Calificación ocupacional de la ocupación principal						
CA01 - Profesional		167	167.00	168	168.00	*****
CA02 - Técnico		224	224.00	225	225.00	*****
CA03 - Operativo		945	945.00	947	947.00	*****
CA04 - No calificado		74	74.00	77	77.00	****
CA05 - Ns/Nc		7	7.00	==RAND.ASSIGN.==		

-----+-----+-----					
47 . Edad (agrupado)					
ED01 - <= 20		39	39.00		39 39.00 **
ED02 - 21 - 30		190	190.00		190 190.00 *****
ED03 - 31 - 40		293	293.00		293 293.00 *****
ED04 - 41 - 50		332	332.00		332 332.00 *****
ED05 - 51 - 60		336	336.00		336 336.00 *****
ED06 - 61 - 70		164	164.00		164 164.00 *****
ED07 - 71 y más		63	63.00		63 63.00 ***
-----+-----+-----					
51 . Ingreso neto laboral total del preceptor sin imputados (agru					
V501 - <= ,00		73	73.00		73 73.00 ****
V502 - ,01 - 50,00		85	85.00		85 85.00 ****
V503 - 50,01 - 100,00		72	72.00		72 72.00 ****
V504 - 100,01 - 150,00		62	62.00		62 62.00 ***
V505 - 150,01 - 200,00		78	78.00		78 78.00 ****
V506 - 200,01 - 250,00		60	60.00		60 60.00 ***
V507 - 250,01 - 308,33		68	68.00		68 68.00 ***
V508 - 308,34 - 400,00		106	106.00		106 106.00 *****
V509 - 400,01 - 450,00		41	41.00		41 41.00 **
V510 - 450,01 - 500,00		68	68.00		68 68.00 ***
V511 - 500,01 - 600,00		88	88.00		88 88.00 *****
V512 - 600,01 - 700,00		57	57.00		57 57.00 ***
V513 - 700,01 - 800,00		67	67.00		67 67.00 ***
V514 - 800,01 - 960,00		67	67.00		67 67.00 ***
V515 - 960,01 - 1063,33		71	71.00		71 71.00 *****
V516 - 1063,34 - 1300,00		80	80.00		80 80.00 ****
V517 - 1300,01 - 1600,00		65	65.00		65 65.00 ***
V518 - 1600,01 - 2000,00		79	79.00		79 79.00 ****
V519 - 2000,01 - 2850,00		60	60.00		60 60.00 ***
V520 - 2850,01+		70	70.00		70 70.00 ***
51_ - missing category		0	0.00		
-----+-----+-----					
52 . Rama de actividad de la ocupación principal resumida					
RA01 - Sector primario		28	28.00		28 28.00 **
RA02 - Fabricación producto		54	54.00		55 55.00 ***
RA03 - Fabricación textil		78	78.00		78 78.00 ****
RA04 - Maquinaria		69	69.00		70 70.00 ***
RA05 - Construcción		193	193.00		193 193.00 *****
RA06 - Comercio		346	346.00		347 347.00 *****
RA07 - Reparación y manteni		85	85.00		85 85.00 ****
RA08 - Hostelería		27	27.00		27 27.00 **
RA09 - Transporte y telecom		113	113.00		114 114.00 *****
RA10 - Finanzas e inmobilia		192	192.00		192 192.00 *****
RA11 - Administración públi		4	4.00		==RAND.ASSIGN.==
RA12 - Enseñanza		57	57.00		57 57.00 ***
RA13 - Salud y servicios so		65	65.00		65 65.00 ***
RA14 - Actividades cultural		34	34.00		34 34.00 **
RA15 - Servicios personales		71	71.00		72 72.00 ****
RA16 - Ns/Nc		1	1.00		==RAND.ASSIGN.==
-----+-----+-----					
53 . Nivel de instrucción					
NI01 - Primario incompleto		146	146.00		146 146.00 *****
NI02 - Primario Completo		336	336.00		336 336.00 *****
NI03 - Secundario incomplet		242	242.00		242 242.00 *****
NI04 - Secundario completo		261	261.00		261 261.00 *****
NI05 - Superior incompleto		36	36.00		36 36.00 **
NI06 - Superior completo		83	83.00		83 83.00 ****
NI07 - Universitario incomp		119	119.00		119 119.00 *****
NI08 - Universitario comple		194	194.00		194 194.00 *****
-----+-----+-----					

EIGENVALUES

COMPUTATIONS PRECISION SUMMARY : TRACE BEFORE DIAGONALISATION.. 6.6250
SUM OF EIGENVALUES..... 6.6250

HISTOGRAM OF THE FIRST 53 EIGENVALUES

NUMBER	EIGENVALUE	PERCENTAGE	CUMULATED PERCENTAGE	
1	0.3564	5.38	5.38	*****
2	0.2661	4.02	9.40	*****
3	0.2343	3.54	12.93	*****
4	0.2133	3.22	16.15	*****
5	0.1878	2.83	18.99	*****
6	0.1781	2.69	21.68	*****
7	0.1620	2.44	24.12	*****
8	0.1553	2.34	26.46	*****
9	0.1538	2.32	28.79	*****
10	0.1490	2.25	31.04	*****
11	0.1466	2.21	33.25	*****
12	0.1436	2.17	35.42	*****
13	0.1427	2.15	37.57	*****
14	0.1417	2.14	39.71	*****
15	0.1390	2.10	41.81	*****
16	0.1375	2.08	43.88	*****
17	0.1338	2.02	45.90	*****
18	0.1322	2.00	47.90	*****
19	0.1308	1.97	49.87	*****
20	0.1304	1.97	51.84	*****
21	0.1293	1.95	53.79	*****
22	0.1280	1.93	55.72	*****
23	0.1264	1.91	57.63	*****
24	0.1256	1.90	59.53	*****
25	0.1245	1.88	61.41	*****
26	0.1235	1.86	63.27	*****
27	0.1215	1.83	65.11	*****
28	0.1203	1.82	66.92	*****
29	0.1185	1.79	68.71	*****
30	0.1171	1.77	70.48	*****
31	0.1165	1.76	72.24	*****
32	0.1146	1.73	73.97	*****
33	0.1129	1.70	75.67	*****
34	0.1115	1.68	77.35	*****
35	0.1078	1.63	78.98	*****
36	0.1068	1.61	80.59	*****
37	0.1067	1.61	82.20	*****
38	0.1041	1.57	83.77	*****
39	0.1016	1.53	85.31	*****
40	0.1004	1.52	86.82	*****
41	0.0969	1.46	88.29	*****
42	0.0925	1.40	89.68	*****
43	0.0917	1.38	91.07	*****
44	0.0888	1.34	92.41	*****
45	0.0836	1.26	93.67	*****
46	0.0793	1.20	94.87	*****
47	0.0674	1.02	95.89	*****
48	0.0632	0.95	96.84	*****
49	0.0606	0.92	97.76	*****
50	0.0494	0.75	98.50	*****
51	0.0437	0.66	99.16	*****
52	0.0315	0.48	99.64	*****
53	0.0241	0.36	100.00	*****

RESEARCH OF IRREGULARITIES (THIRD DIFFERENCES)

IRREGULARITY	IRREGULARITY	
BETWEEN	VALUE	
1 -- 2	-47.68	*****
4 -- 5	-22.26	*****
2 -- 3	-15.29	*****
49 -- 50	-10.19	*****
7 -- 8	-8.63	*****
46 -- 47	-8.33	*****
41 -- 42	-5.74	*****
6 -- 7	-4.20	*****
40 -- 41	-3.69	*****
16 -- 17	-3.51	****
37 -- 38	-3.43	****
9 -- 10	-3.19	****
34 -- 35	-2.55	***
31 -- 32	-2.12	***
11 -- 12	-2.07	***
14 -- 15	-1.74	**
18 -- 19	-1.72	**
26 -- 27	-1.34	**
22 -- 23	-1.11	**
44 -- 45	-0.02	*

RESEARCH OF IRREGULARITIES (SECOND DIFFERENCES)

IRREGULARITY	IRREGULARITY	
BETWEEN	VALUE	
1 -- 2	58.46	*****
4 -- 5	15.85	*****
2 -- 3	10.78	*****
6 -- 7	9.42	*****
46 -- 47	7.67	*****
49 -- 50	5.49	****
7 -- 8	5.23	****
41 -- 42	3.59	****
34 -- 35	2.78	***
9 -- 10	2.52	***
16 -- 17	2.16	**
11 -- 12	2.06	**
47 -- 48	1.59	**
38 -- 39	1.37	**
14 -- 15	1.26	**
18 -- 19	1.01	*
35 -- 36	0.86	*
22 -- 23	0.86	*
29 -- 30	0.80	*
44 -- 45	0.80	*
26 -- 27	0.77	*
28 -- 29	0.44	*
32 -- 33	0.26	*
31 -- 32	0.18	*
17 -- 18	0.14	*
37 -- 38	0.13	*
24 -- 25	0.04	*

LOADINGS, CONTRIBUTIONS AND SQUARED COSINES OF ACTIVE CATEGORIES

AXES 1 TO 2

CATEGORIES			LOADINGS					CONTRIBUTIONS					SQUARED COSINES				
IDEN	LABEL	REL. WT. DISTO	1	2	0	0	0	1	2	0	0	0	1	2	0	0	0
8 . ¿La semana pasada tenía ...																	
CP01	- un solo empleo/ ocup	11.43 0.09	-0.05	0.00	0.00	0.00	0.00	0.1	0.0	0.0	0.0	0.0	0.02	0.00	0.00	0.00	0.00
CP02	- 2 o más empleos	1.07 10.71	0.51	0.03	0.00	0.00	0.00	0.8	0.0	0.0	0.0	0.0	0.02	0.00	0.00	0.00	0.00
CUMULATED CONTRIBUTION = 0.8 0.0 0.0 0.0 0.0																	
32 . Relación de parentesco																	
CP01	- Jefe/a	7.53 0.66	0.11	0.46	0.00	0.00	0.00	0.3	6.1	0.0	0.0	0.0	0.02	0.32	0.00	0.00	0.00
CP02	- Cónyuge/pareja	3.07 3.07	0.01	-0.59	0.00	0.00	0.00	0.0	4.0	0.0	0.0	0.0	0.00	0.11	0.00	0.00	0.00
CP03	- Hijo/hijastro/a	1.54 7.10	-0.38	-1.03	0.00	0.00	0.00	0.6	6.1	0.0	0.0	0.0	0.02	0.15	0.00	0.00	0.00
CP04	- Otro parentesco	0.35 34.42	-0.80	-0.24	0.00	0.00	0.00	0.6	0.1	0.0	0.0	0.0	0.02	0.00	0.00	0.00	0.00
CUMULATED CONTRIBUTION = 1.5 16.3 0.0 0.0 0.0																	
33 . Sexo																	
CP01	- Varón	7.73 0.62	-0.04	0.48	0.00	0.00	0.00	0.0	6.8	0.0	0.0	0.0	0.00	0.38	0.00	0.00	0.00
CP02	- Mujer	4.77 1.62	0.07	-0.78	0.00	0.00	0.00	0.1	11.0	0.0	0.0	0.0	0.00	0.38	0.00	0.00	0.00
CUMULATED CONTRIBUTION = 0.1 17.9 0.0 0.0 0.0																	
39 . Calificación ocupacional de la ocupación principal																	
CA01	- Profesional	1.48 7.43	2.16	0.28	0.00	0.00	0.00	19.4	0.4	0.0	0.0	0.0	0.63	0.01	0.00	0.00	0.00
CA02	- Técnico	1.98 5.30	0.57	-1.16	0.00	0.00	0.00	1.8	10.0	0.0	0.0	0.0	0.06	0.25	0.00	0.00	0.00
CA03	- Operativo	8.35 0.50	-0.45	0.25	0.00	0.00	0.00	4.8	2.0	0.0	0.0	0.0	0.41	0.13	0.00	0.00	0.00
CA04	- No calificado	0.68 17.40	-0.82	-0.33	0.00	0.00	0.00	1.3	0.3	0.0	0.0	0.0	0.04	0.01	0.00	0.00	0.00
CUMULATED CONTRIBUTION = 27.3 12.7 0.0 0.0 0.0																	
47 . Edad (agrupado)																	
ED01	- <= 20	0.34 35.33	-1.02	-1.60	0.00	0.00	0.00	1.0	3.3	0.0	0.0	0.0	0.03	0.07	0.00	0.00	0.00
ED02	- 21 - 30	1.68 6.46	-0.21	-0.68	0.00	0.00	0.00	0.2	2.9	0.0	0.0	0.0	0.01	0.07	0.00	0.00	0.00
ED03	- 31 - 40	2.58 3.84	0.16	-0.10	0.00	0.00	0.00	0.2	0.1	0.0	0.0	0.0	0.01	0.00	0.00	0.00	0.00
ED04	- 41 - 50	2.93 3.27	0.02	0.16	0.00	0.00	0.00	0.0	0.3	0.0	0.0	0.0	0.00	0.01	0.00	0.00	0.00
ED05	- 51 - 60	2.96 3.22	0.09	0.31	0.00	0.00	0.00	0.1	1.1	0.0	0.0	0.0	0.00	0.03	0.00	0.00	0.00
ED06	- 61 - 70	1.45 7.64	0.06	0.46	0.00	0.00	0.00	0.0	1.1	0.0	0.0	0.0	0.00	0.03	0.00	0.00	0.00
ED07	- 71 y más	0.56 21.49	-0.24	-0.15	0.00	0.00	0.00	0.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
CUMULATED CONTRIBUTION = 1.6 8.9 0.0 0.0 0.0																	
51 . Ingreso neto laboral total del preceptor sin imputados (agrupado)																	
V501	- <= ,00	0.64 18.41	-0.35	-0.78	0.00	0.00	0.00	0.2	1.5	0.0	0.0	0.0	0.01	0.03	0.00	0.00	0.00
V502	- ,01 - 50,00	0.75 15.67	-0.64	-1.01	0.00	0.00	0.00	0.9	2.9	0.0	0.0	0.0	0.03	0.07	0.00	0.00	0.00
V503	- 50,01 - 100,00	0.64 18.68	-0.38	-0.89	0.00	0.00	0.00	0.3	1.9	0.0	0.0	0.0	0.01	0.04	0.00	0.00	0.00
V504	- 100,01 - 150,00	0.55 21.85	-0.53	-0.44	0.00	0.00	0.00	0.4	0.4	0.0	0.0	0.0	0.01	0.01	0.00	0.00	0.00
V505	- 150,01 - 200,00	0.69 17.17	-0.29	-0.43	0.00	0.00	0.00	0.2	0.5	0.0	0.0	0.0	0.00	0.01	0.00	0.00	0.00
V506	- 200,01 - 250,00	0.53 22.62	-0.40	-0.15	0.00	0.00	0.00	0.2	0.0	0.0	0.0	0.0	0.01	0.00	0.00	0.00	0.00
V507	- 250,01 - 308,33	0.60 19.84	-0.61	0.09	0.00	0.00	0.00	0.6	0.0	0.0	0.0	0.0	0.02	0.00	0.00	0.00	0.00
V508	- 308,34 - 400,00	0.94 12.37	-0.51	0.19	0.00	0.00	0.00	0.7	0.1	0.0	0.0	0.0	0.02	0.00	0.00	0.00	0.00
V509	- 400,01 - 450,00	0.36 33.56	-0.59	0.21	0.00	0.00	0.00	0.4	0.1	0.0	0.0	0.0	0.01	0.00	0.00	0.00	0.00
V510	- 450,01 - 500,00	0.60 19.84	-0.15	-0.03	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
V511	- 500,01 - 600,00	0.78 15.10	-0.19	0.28	0.00	0.00	0.00	0.1	0.2	0.0	0.0	0.0	0.00	0.01	0.00	0.00	0.00
V512	- 600,01 - 700,00	0.50 23.86	-0.29	0.49	0.00	0.00	0.00	0.1	0.4	0.0	0.0	0.0	0.00	0.01	0.00	0.00	0.00
V513	- 700,01 - 800,00	0.59 20.15	-0.19	0.37	0.00	0.00	0.00	0.1	0.3	0.0	0.0	0.0	0.00	0.01	0.00	0.00	0.00
V514	- 800,01 - 960,00	0.59 20.15	0.04	0.38	0.00	0.00	0.00	0.0	0.3	0.0	0.0	0.0	0.00	0.01	0.00	0.00	0.00
V515	- 960,01 - 1063,33	0.63 18.96	0.27	0.28	0.00	0.00	0.00	0.1	0.2	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
V516	- 1063,34 - 1300,00	0.71 16.71	0.22	0.48	0.00	0.00	0.00	0.1	0.6	0.0	0.0	0.0	0.00	0.01	0.00	0.00	0.00
V517	- 1300,01 - 1600,00	0.57 20.80	0.55	0.11	0.00	0.00	0.00	0.5	0.0	0.0	0.0	0.0	0.01	0.00	0.00	0.00	0.00
V518	- 1600,01 - 2000,00	0.70 16.94	0.99	0.11	0.00	0.00	0.00	1.9	0.0	0.0	0.0	0.0	0.06	0.00	0.00	0.00	0.00
V519	- 2000,01 - 2850,00	0.53 22.62	1.46	0.47	0.00	0.00	0.00	3.2	0.4	0.0	0.0	0.0	0.09	0.01	0.00	0.00	0.00
V520	- 2850,01+	0.62 19.24	1.75	0.55	0.00	0.00	0.00	5.3	0.7	0.0	0.0	0.0	0.16	0.02	0.00	0.00	0.00
CUMULATED CONTRIBUTION = 15.2 10.7 0.0 0.0 0.0																	
52 . Rama de actividad de la ocupación principal resumida																	
RA01	- Sector primario	0.25 49.61	-0.89	-0.11	0.00	0.00	0.00	0.5	0.0	0.0	0.0	0.0	0.02	0.00	0.00	0.00	0.00
RA02	- Fabricación producto	0.49 24.76	-0.26	0.24	0.00	0.00	0.00	0.1	0.1	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
RA03	- Fabricación textil	0.69 17.17	-0.51	-0.43	0.00	0.00	0.00	0.5	0.5	0.0	0.0	0.0	0.01	0.01	0.00	0.00	0.00
RA04	- Maquinaria	0.62 19.24	-0.23	-0.08	0.00	0.00	0.00	0.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
RA05	- Construcción	1.70 6.34	-0.67	0.89	0.00	0.00	0.00	2.2	5.1	0.0	0.0	0.0	0.07	0.12	0.00	0.00	0.00
RA06	- Comercio	3.06 3.08	-0.49	-0.14	0.00	0.00	0.00	2.0	0.2	0.0	0.0	0.0	0.08	0.01	0.00	0.00	0.00
RA07	- Reparación y manteni	0.75 15.67	-0.33	0.53	0.00	0.00	0.00	0.2	0.8	0.0	0.0	0.0	0.01	0.02	0.00	0.00	0.00
RA08	- Hostelería	0.24 51.48	-0.48	-0.59	0.00	0.00	0.00	0.2	0.3	0.0	0.0	0.0	0.00	0.01	0.00	0.00	0.00
RA09	- Transporte y telecom	1.01 11.43	-0.33	0.83	0.00	0.00	0.00	0.3	2.6	0.0	0.0	0.0	0.01	0.06	0.00	0.00	0.00
RA10	- Finanzas e inmobilia	1.69 6.38	1.56	0.07	0.00	0.00	0.00	11.5	0.0	0.0	0.0	0.0	0.38	0.00	0.00	0.00	0.00
RA12	- Enseñanza	0.50 23.86	0.33	-2.61	0.00	0.00	0.00	0.2	12.8	0.0	0.0	0.0	0.00	0.28	0.00	0.00	0.00
RA13	- Salud y servicios so	0.57 20.80	2.19	-0.07	0.00	0.00	0.00	7.7	0.0	0.0	0.0	0.0	0.23	0.00	0.00	0.00	0.00
RA14	- Actividades cultural	0.30 40.68	0.82	-0.62	0.00	0.00	0.00	0.6	0.4	0.0	0.0	0.0	0.02	0.01	0.00	0.00	0.00
RA15	- Servicios personales	0.64 18.68	-0.23	-0.78	0.00	0.00	0.00	0.1	1.4	0.0	0.0	0.0	0.00	0.03	0.00	0.00	0.00
CUMULATED CONTRIBUTION = 26.2 24.4 0.0 0.0 0.0																	

CATEGORIES			LOADINGS					CONTRIBUTIONS					SQUARED COSINES				
IDEN - LABEL	REL. WT.	DISTO	1	2	0	0	0	1	2	0	0	0	1	2	0	0	0
53 . Nivel de instrucción																	
NI01 - Primario incompleto	1.29	8.71	-0.72	0.29	0.00	0.00	0.00	1.9	0.4	0.0	0.0	0.0	0.06	0.01	0.00	0.00	0.00
NI02 - Primario Completo	2.96	3.22	-0.59	0.39	0.00	0.00	0.00	2.9	1.7	0.0	0.0	0.0	0.11	0.05	0.00	0.00	0.00
NI03 - Secundario incompleto	2.13	4.86	-0.44	0.16	0.00	0.00	0.00	1.2	0.2	0.0	0.0	0.0	0.04	0.01	0.00	0.00	0.00
NI04 - Secundario completo	2.30	4.43	-0.19	-0.12	0.00	0.00	0.00	0.2	0.1	0.0	0.0	0.0	0.01	0.00	0.00	0.00	0.00
NI05 - Superior incompleto	0.32	38.36	-0.17	-1.37	0.00	0.00	0.00	0.0	2.2	0.0	0.0	0.0	0.00	0.05	0.00	0.00	0.00
NI06 - Superior completo	0.73	16.07	0.51	-1.10	0.00	0.00	0.00	0.5	3.3	0.0	0.0	0.0	0.02	0.07	0.00	0.00	0.00
NI07 - Universitario incompleto	1.05	10.91	0.19	-0.55	0.00	0.00	0.00	0.1	1.2	0.0	0.0	0.0	0.00	0.03	0.00	0.00	0.00
NI08 - Universitario completo	1.71	6.30	2.06	0.13	0.00	0.00	0.00	20.4	0.1	0.0	0.0	0.0	0.67	0.00	0.00	0.00	0.00
CUMULATED CONTRIBUTION =									27.2	9.3	0.0	0.0	0.0				

LOADINGS AND TEST-VALUES OF CATEGORIES

AXES 1 TO 2

CATEGORIES			TEST-VALUES					LOADINGS					DISTO.				
IDEN - LABEL	COUNT	ABS.WT	1	2	0	0	0	1	2	0	0	0	1	2	0	0	0
8 . ¿La semana pasada tenía ...																	
CP01 - un solo empleo/ ocup	1296	1296.00	-5.8	-0.3	0.0	0.0	0.0	-0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09
CP02 - 2 o más empleos	121	121.00	5.8	0.3	0.0	0.0	0.0	0.51	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.71
32 . Relación de parentesco																	
CP01 - Jefe/a	854	854.00	5.1	21.5	0.0	0.0	0.0	0.11	0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.66
CP02 - Cónyuge/pareja	348	348.00	0.3	-12.7	0.0	0.0	0.0	0.01	-0.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.07
CP03 - Hijo/hijastro/a	175	175.00	-5.4	-14.5	0.0	0.0	0.0	-0.38	-1.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.10
CP04 - Otro parentesco	40	40.00	-5.1	-1.5	0.0	0.0	0.0	-0.80	-0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	34.42
33 . Sexo																	
CP01 - Varón	876	876.00	-1.9	23.2	0.0	0.0	0.0	-0.04	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.62
CP02 - Mujer	541	541.00	1.9	-23.2	0.0	0.0	0.0	0.07	-0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.62
39 . Calificación ocupacional de la ocupación principal																	
CA01 - Profesional	167	167.00	29.9	4.0	0.0	0.0	0.0	2.17	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.49
CA02 - Técnico	224	224.00	9.4	-19.0	0.0	0.0	0.0	0.58	-1.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.33
CA03 - Operativo	945	945.00	-24.2	13.4	0.0	0.0	0.0	-0.45	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50
CA04 - No calificado	74	74.00	-7.4	-3.0	0.0	0.0	0.0	-0.83	-0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	18.15
CA05 - Ns/Nc	7	7.00	-0.3	-0.2	0.0	0.0	0.0	-0.11	-0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	201.43
47 . Edad (agrupado)																	
ED01 - <= 20	39	39.00	-6.5	-10.1	0.0	0.0	0.0	-1.02	-1.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	35.33
ED02 - 21 - 30	190	190.00	-3.1	-10.1	0.0	0.0	0.0	-0.21	-0.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.46
ED03 - 31 - 40	293	293.00	3.1	-1.9	0.0	0.0	0.0	0.16	-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.84
ED04 - 41 - 50	332	332.00	0.5	3.2	0.0	0.0	0.0	0.02	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.27
ED05 - 51 - 60	336	336.00	1.8	6.5	0.0	0.0	0.0	0.09	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.22
ED06 - 61 - 70	164	164.00	0.9	6.2	0.0	0.0	0.0	0.06	0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.64
ED07 - 71 y más	63	63.00	-2.0	-1.3	0.0	0.0	0.0	-0.24	-0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	21.49
51 . Ingreso neto laboral total del preceptor sin imputados (agrupado)																	
V501 - <= ,00	73	73.00	-3.1	-6.9	0.0	0.0	0.0	-0.35	-0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	18.41
V502 - ,01 - 50,00	85	85.00	-6.1	-9.6	0.0	0.0	0.0	-0.64	-1.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	15.67
V503 - 50,01 - 100,00	72	72.00	-3.3	-7.7	0.0	0.0	0.0	-0.38	-0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	18.68
V504 - 100,01 - 150,00	62	62.00	-4.2	-3.5	0.0	0.0	0.0	-0.53	-0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	21.85
V505 - 150,01 - 200,00	78	78.00	-2.6	-3.9	0.0	0.0	0.0	-0.29	-0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	17.17
V506 - 200,01 - 250,00	60	60.00	-3.2	-1.2	0.0	0.0	0.0	-0.40	-0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	22.62
V507 - 250,01 - 308,33	68	68.00	-5.2	0.8	0.0	0.0	0.0	-0.61	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	19.84
V508 - 308,34 - 400,00	106	106.00	-5.5	2.1	0.0	0.0	0.0	-0.51	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	12.37
V509 - 400,01 - 450,00	41	41.00	-3.9	1.4	0.0	0.0	0.0	-0.59	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	33.56
V510 - 450,01 - 500,00	68	68.00	-1.3	-0.2	0.0	0.0	0.0	-0.15	-0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	19.84
V511 - 500,01 - 600,00	88	88.00	-1.8	2.7	0.0	0.0	0.0	-0.19	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	15.10
V512 - 600,01 - 700,00	57	57.00	-2.2	3.8	0.0	0.0	0.0	-0.29	0.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00	23.86
V513 - 700,01 - 800,00	67	67.00	-1.6	3.1	0.0	0.0	0.0	-0.19	0.37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.15
V514 - 800,01 - 960,00	67	67.00	0.3	3.2	0.0	0.0	0.0	0.04	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.15
V515 - 960,01 - 1063,33	71	71.00	2.3	2.4	0.0	0.0	0.0	0.27	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	18.96
V516 - 1063,34 - 1300,00	80	80.00	2.1	4.4	0.0	0.0	0.0	0.22	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	16.71
V517 - 1300,01 - 1600,00	65	65.00	4.5	0.9	0.0	0.0	0.0	0.55	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.80
V518 - 1600,01 - 2000,00	79	79.00	9.1	1.0	0.0	0.0	0.0	0.99	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	16.94
V519 - 2000,01 - 2850,00	60	60.00	11.5	3.7	0.0	0.0	0.0	1.46	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	22.62
V520 - 2850,01+	70	70.00	15.0	4.7	0.0	0.0	0.0	1.75	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	19.24
51_ - missing category	0	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+														
	52 . Rama de actividad de la ocupación principal resumida													
	RA01 - Sector primario	28	28.00		-4.7	-0.6	0.0	0.0	0.0		-0.89	-0.11	0.00	0.00
	RA02 - Fabricación producto	54	54.00		-2.0	1.9	0.0	0.0	0.0		-0.26	0.25	0.00	0.00
	RA03 - Fabricación textil	78	78.00		-4.6	-3.9	0.0	0.0	0.0		-0.51	-0.43	0.00	0.00
	RA04 - Maquinaria	69	69.00		-1.9	-0.7	0.0	0.0	0.0		-0.22	-0.09	0.00	0.00
	RA05 - Construcción	193	193.00		-10.1	13.3	0.0	0.0	0.0		-0.67	0.89	0.00	0.00
	RA06 - Comercio	346	346.00		-10.5	-3.1	0.0	0.0	0.0		-0.49	-0.14	0.00	0.00
	RA07 - Reparación y manteni	85	85.00		-3.1	5.0	0.0	0.0	0.0		-0.33	0.53	0.00	0.00
	RA08 - Hostelería	27	27.00		-2.5	-3.1	0.0	0.0	0.0		-0.48	-0.59	0.00	0.00
	RA09 - Transporte y telecom	113	113.00		-3.8	9.2	0.0	0.0	0.0		-0.34	0.83	0.00	0.00
	RA10 - Finanzas e inmobilia	192	192.00		23.2	1.0	0.0	0.0	0.0		1.56	0.07	0.00	0.00
	RA11 - Administración públi	4	4.00		0.9	0.2	0.0	0.0	0.0		0.43	0.11	0.00	0.00
	RA12 - Enseñanza	57	57.00		2.5	-20.1	0.0	0.0	0.0		0.33	-2.61	0.00	0.00
	RA13 - Salud y servicios so	65	65.00		18.1	-0.6	0.0	0.0	0.0		2.19	-0.07	0.00	0.00
	RA14 - Actividades cultural	34	34.00		4.8	-3.6	0.0	0.0	0.0		0.82	-0.62	0.00	0.00
	RA15 - Servicios personales	71	71.00		-2.0	-6.7	0.0	0.0	0.0		-0.23	-0.77	0.00	0.00
	RA16 - Ns/Nc	1	1.00		0.0	0.1	0.0	0.0	0.0		-0.02	0.06	0.00	0.00
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+														
	53 . Nivel de instrucción													
	NI01 - Primario incompleto	146	146.00		-9.2	3.7	0.0	0.0	0.0		-0.72	0.29	0.00	0.00
	NI02 - Primario Completo	336	336.00		-12.3	8.1	0.0	0.0	0.0		-0.59	0.39	0.00	0.00
	NI03 - Secundario incomplet	242	242.00		-7.5	2.8	0.0	0.0	0.0		-0.44	0.16	0.00	0.00
	NI04 - Secundario completo	261	261.00		-3.3	-2.2	0.0	0.0	0.0		-0.19	-0.12	0.00	0.00
	NI05 - Superior incompleto	36	36.00		-1.0	-8.3	0.0	0.0	0.0		-0.17	-1.37	0.00	0.00
	NI06 - Superior completo	83	83.00		4.8	-10.3	0.0	0.0	0.0		0.51	-1.10	0.00	0.00
	NI07 - Universitario incomp	119	119.00		2.2	-6.2	0.0	0.0	0.0		0.19	-0.55	0.00	0.00
	NI08 - Universitario comple	194	194.00		30.9	2.0	0.0	0.0	0.0		2.06	0.13	0.00	0.00
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+														
	17 . Marca del ingreso laboral total del perceptor													
	M_01 - ingreso imputado	0	0.00		0.0	0.0	0.0	0.0	0.0		0.00	0.00	0.00	0.00
	M_02 - no perceptor	3	3.00		-1.1	-0.7	0.0	0.0	0.0		-0.64	-0.40	0.00	0.00
	M_03 - ingreso sin imputar	1414	1414.00		1.1	0.7	0.0	0.0	0.0		0.00	0.00	0.00	0.00
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+														

CORRELATIONS BETWEEN CONTINUOUS VARIABLES AND FACTORS

AXES 1 A 2

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+														
	VARIABLES				SUMMARY STATISTICS					CORRELATIONS				
	-----+-----+-----+-----				-----+-----+-----+-----+-----					-----+-----+-----+-----+-----				
	NUM . (IDEN)	SHORT	LABEL		COUNT	ABS.WT	MEAN	ST.DEV.		1	2	0	0	0
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+														
	16 . (INGL)	Ingreso	neto laboral		1417	1417.00	862.22	1115.75		0.58	0.26	0.00	0.00	0.00
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+														